```
                 GenCore version 5.1.4_p5_4578
       Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:          May 16, 2003, 12:14:56 ; Search time 113 Seconds
                                          (without alignments)
                                          5044.656 Million cell updates/sec

Title:           US-10-008-355-1
Perfect score:   3820
Sequence:        1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters:      1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110337_9308/app_query.fasta_1.2311
-DB=A_Geneseq_101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_153_@runat_16052003_110337_9308 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       A_Geneseq_101002:*
                 1:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
                 2:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
                 3:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
                 4:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
                 5:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
                 6:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
                 7:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
                 8:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
                 9:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
                 10:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
                 11:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
                 12:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
                 13:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
                 14:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
                 15:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
                 16:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
                 17:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
                 18:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
                 19:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
                 20:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
                 21:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
                 22:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
                 23:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3719	97.4	712	23	AAO15205	Porphyromonas ging
	2	274	7.2	52	23	AAO15206	Porphyromonas ging
	3	174	4.6	1017	22	AAB59813	TutD protein #4.
	4	174	4.6	1615	22	AAB59826	Protein #3 encoded
c	5	154.5	3.9	1592	22	AAB59827	Protein #4 encoded
c	6	153.5	3.9	999	22	AAB59817	TutD protein #8.
	7	153.5	4.0	1296	23	ABG66702	Human novel polype
	8	153.5	4.0	1296	23	ABG66756	Human novel polype
	9	148.5	3.9	1528	18	AAW20056	C. elegans UNC-53
	10	148.5	3.9	1583	18	AAW20057	C. elegans UNC-53
	11	148	3.9	2570	22	ABG06375	Novel human diagno
	12	146.5	3.8	1368	23	ABP27518	Streptococcus poly
	13	141.5	3.7	1795	22	ABB69806	Drosophila melanog
c	14	138	3.5	430	20	AAY31745	Mycobacterium tube
c	15	138	3.5	684	22	ABB61705	Drosophila melanog
	16	135	3.5	1605	22	AAB59824	Protein #1 encoded
	17	134.5	3.5	1022	22	AAB59814	TutD protein #5.
	18	133	3.5	883	22	AAU37789	Streptococcus pneu
c	19	132	3.3	571	20	AAY36845	Protein involved i
	20	130	3.4	883	20	AAY08339	S. pneumoniae valS
	21	130	3.4	883	21	AAY90514	Streptococcus pneu
	22	130	3.4	883	23	AAU97883	Streptococcus pneu
	23	130	3.4	1061	20	AAW87504	Human N-methyl-D-a
	24	127.5	3.3	1194	18	AAW21725	Modified streptoki
c	25	127	3.2	388	20	AAY04998	Mycobacterium spec
	26	127	3.3	883	22	AAU38091	Streptococcus pneu
	27	127	3.3	1592	22	AAB59827	Protein #4 encoded
	28	126.5	3.3	1370	23	ABP27517	Streptococcus poly
	29	126.5	3.3	2870	21	AAY95559	Caenorhabditis ele
	30	126.5	3.3	3178	21	AAY95556	Caenorhabditis ele
	31	124	3.2	571	22	AAB27242	Human EXMAD-20 SEQ
	32	123.5	3.2	713	20	AAY43380	M. catarrhalis str
	33	122.5	3.2	26	23	AAO15221	Porphyromonas ging
	34	122.5	3.2	711	17	AAR88649	Neisseria meningit
	35	122.5	3.2	800	18	AAW21723	Plasminogen-bindin
	36	122.5	3.2	813	18	AAW21728	Wild type plasmino
	37	122.5	3.2	1181	18	AAW21727	Streptokinase/malt
	38	122.5	3.2	1194	18	AAW21724	Modified streptoki
	39	122.5	3.2	1194	18	AAW21726	Streptokinase/malt
c	40	122	3.1	677	22	AAU57954	Propionibacterium
	41	121.5	3.2	746	22	ABB59201	Drosophila melanog
	42	121	3.2	2432	21	AAY85565	Human homologue of
	43	120	3.1	999	22	AAB59817	TutD protein #8.
c	44	120	3.0	2467	23	ABB91303	Herbicidally activ
	45	119.5	3.1	2042	19	AAW56319	Haemophilus paraga

ALIGNMENTS

```
RESULT 1
AAO15205
ID   AAO15205 standard; Protein; 712 AA.
XX
AC   AAO15205;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7).
XX
KW   Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
KW   DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW   periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200238742-A2.
XX
PD   16-MAY-2002.
```

```
XX
PF   08-NOV-2001; 2001WO-US46782.
XX
PR   08-NOV-2000; 2000US-246827P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,  Potempa JS,  Banbula A,  Bugno M;
XX
DR   WPI; 2002-490075/52.
DR   N-PSDB; AAL43635.
XX
PT   Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT   the dipeptidylpeptidase for protecting an animal from periodontal
PT   disease caused by Porphyromonas gingivalis  -
XX
PS   Claim 7; Fig 4; 65pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC   enzyme has amidolytic activity for cleavage of a peptide bond between the
CC   second and third amino acids from the N-terminal end of a target peptide.
CC   The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC   substituent on the alpha-carbon atom of the second amino acid from the
CC   N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC   useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC   reducing the growth of a bacterium and protecting an animal from a
CC   periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC   or periodontitis). The present amino acid sequence represents the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the
CC   invention.
XX
SQ   Sequence   712 AA;

Alignment Scores:
Pred. No.:                0            Length:         712
Score:                    3719.00      Matches:        712
Percent Similarity:       100.00%      Conservative:   0
Best Local Similarity:    100.00%      Mismatches:     0
Query Match:              97.36%       Indels:         0
DB:                       23           Gaps:           0

US-10-008-355-1 (1-2139) x AAO15205 (1-712)

Qy        1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20

Qy       61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAsp 40

Qy      121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60

Qy      181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 SerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80

Qy      241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       81 GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100

Qy      301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuPro 120

Qy      361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140

Qy      421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160

Qy      481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180

Qy      541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200

Qy      601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro 220

Qy      661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240

Qy      721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln 260

Qy      781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280

Qy      841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300

Qy      901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320

Qy      961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340

Qy     1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360

Qy     1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380

Qy     1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400

Qy     1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420

Qy     1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 AspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSer 440

Qy     1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAsp 460

Qy     1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480

Qy     1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500

Qy     1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520

Qy    1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540

Qy    1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560

Qy    1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580

Qy    1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600

Qy    1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620

Qy    1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640

Qy    1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660

Qy    1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680

Qy    2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGly 700

Qy    2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATC 2136
           ||||||||||||||||||||||||||||||||||||
Db     701 GlnCysProArgLeuIleGlnGluLeuLysLeuIle 712

RESULT 2
AAO15206
ID   AAO15206 standard; Protein; 52 AA.
XX
AC   AAO15206;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) C-terminal region.
XX
KW   Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
KW   DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW   periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200238742-A2.
XX
PD   16-MAY-2002.
XX
PF   08-NOV-2001; 2001WO-US46782.
XX
PR   08-NOV-2000; 2000US-246827P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,  Potempa JS,  Banbula A,  Bugno M;
XX
DR   WPI; 2002-490075/52.
XX
PT   Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT   the dipeptidylpeptidase for protecting an animal from periodontal
PT   disease caused by Porphyromonas gingivalis  -
XX
PS   Example 6; Fig 5; 65pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC   enzyme has amidolytic activity for cleavage of a peptide bond between the
CC   second and third amino acids from the N-terminal end of a target peptide.
CC   The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC   substituent on the alpha-carbon atom of the second amino acid from the
CC   N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC   useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC   reducing the growth of a bacterium and protecting an animal from a
CC   periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC   or periodontitis). The present amino acid sequence represents the
CC   C-terminal region of the Porphyromonas gingivalis dipeptidylpeptidase-7
CC   (DPP-7) enzyme.
XX
SQ   Sequence   52 AA;

Alignment Scores:
Pred. No.:              8.98e-18       Length:         52
Score:                  274.00         Matches:        52
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            7.17%          Indels:         0
DB:                     23             Gaps:           0

US-10-008-355-1 (1-2139) x AAO15206 (1-52)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCT 1989
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAla 20

Qy    1990 TTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGC 2049
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArg 40

Qy    2050 ACAATCAGCGTGGACATCCGCTACGTTCTCTTCATG 2085
           ||||||||||||||||||||||||||||||||||||
Db      41 ThrIleSerValAspIleArgTyrValLeuPheMet 52

RESULT 3
AAB59813
ID   AAB59813 standard; Protein; 1017 AA.
XX
AC   AAB59813;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #4.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
```

```
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   1017 AA;

Alignment Scores:
Pred. No.:              3.52e-07      Length:        1017
Score:                  174.00        Matches:       176
Percent Similarity:     34.00%        Conservative:  78
Best Local Similarity:  23.56%        Mismatches:    275
Query Match:            4.55%         Indels:        218
DB:                     22            Gaps:          40

US-10-008-355-1 (1-2139) x AAB59813 (1-1017)

Qy     65 CCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAA 124
          |||||||||         |||      |||      :::   ||||||:::
Db    111 ProLysProArgProThrCysArgProSerProGlyThrAlaArgArgValSerThrThr 130

Qy    125 TGCGTGAGCTCGGCTTTACGCTCCCGTTGG------------------------------ 154
                   |||            ||||||
Db    131 SerProArgSerThrThrGlyArgArgTrpSerSerProAlaArgArgSerAlaGlyArg 150

Qy    155 ------------------ATTCGCTCTACAGTTTCGACAAGCCGTCCATTG------CCA 190
                               ||||||:::      ||||||||||||:::
Db    151 AlaGlyArgAlaGlyCysAlaArgSerSerArgLysThrSerArgProIleArgSerAla 170

Qy    191 ATGCCGTGGTTATCT------------TCGGTGGCGGATGTACCG--------------- 223
             |||      |||            ||||||:::      |||
Db    171 ArgProSerCysSerLysSerProThrSerValSerAlaPheProProSerProAlaArg 190

Qy    224 GTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTA 283
             |||:::         |||      ||||||||::::::          :::
Db    191 AlaSerArgThrArgCysArgArgAsnSerLeuProSerSerValThrArgSerSerAla 210

Qy    284 TCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGA 343
          :::||||||                  |||   |||
Db    211 ThrArgAlaAlaThrProArgArgLysThrProCys------------------------ 222

Qy    344 TGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGG 403
                                                    |||   |||:::   |||
Db    223 ------------------------------------------CysGlyArgThrThrArg 228

Qy    404 TAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCA 463
                :::      ::::::|||||||||   :::      ||||||:::
Db    229 ProProSerSerThrArgAsnSerSerArgAlaThrTrpMetArgTrpAsnSerSerArg 248

Qy    464 AAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCT 523
                   |||:::            |||      :::      |||      :::|||
Db    249 TrpAsnValArgPheProSerMetAlaProAlaSerArgAlaProThrAlaLysSerSer 268

Qy    524 GCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCA 583
                         :::|||   :::::::::   |||::::::   |||   . :::
Db    269 Arg------GlyArgThrIleCysSerSerSerProSerAlaAlaProThrProArgAla 286

Qy    584 AGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGG 643
          ||||||                     ||||||         :::      :::   :::
Db    287 ArgThr---------------------ProAlaThrThrProThrProSerSerArgGln 299

Qy    644 ACAACTGGATGTGGCCGC------------------------GTCACACGGGCGACT--- 676
             :::|||      |||                        :::   |||   |||
Db    300 ProSerGlySerAlaArgProSerProProSerSerSerAlaIleProArgArgThrAla 319

Qy    677 ------------------------------TCAGCGTATTCCGCGTGTATGCCGGTGCCG 706
                                        ||||||      :::   :::
Db    320 ArgArgArgCysAlaGlyPheSerSerAlaSerAlaThrAspSerAlaIleArgArgSer 339

Qy    707 ACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTG 766
          :::|||   |||         :::|||         |||   |||      |||:::
Db    340 SerThrThrArgSerAlaArgSerArgArgAsnThrProSerSerAlaSerThrAlaThr 359

Qy    767 CCGTATCCATGCAAG---------------GCTACAAGGCTGACGACTATGCCATGACCA 811
                |||   :::               :::|||               |||
Db    360 AlaProProThrArgLysProThrThrGlySerThrCysCysAlaCysArgProAlaSer 379

Qy    812 TCGGTTTCC---------------------CGGGCAGTACGGATCGCTACCTCACTTCTT 850
          :::|||:::                     |||   |||         :::|||
Db    380 ThrValAlaAlaArgArgLysLysProValArgLysValAlaAlaGlnSerSerArgPro 399

Qy    851 GGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCA 910
                ||||||      |||   |||   |||
Db    400 SerCysTrpLysSerArgSerMetThrAlaThr--------------------------- 410

Qy    911 AGCAAGGCATCTGGAAGGAAGCCATGAGCG---CAGATCAGGCTACCCGTA------TCA 961
                   :::||||||   |||               |||   ||||||
Db    411 ---------ThrGlyArgThrProThrCysAsnSerAlaArgArgProValIleSerArg 427

Qy    962 AATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACC 1021
                |||:::   |||                  ||||||:::            :::
Db    428 ArgSerProSerArgMet---------------PheGlyArgLeu------------Ser 438

Qy   1022 GCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAG 1081
          |||   |||:::   :::            ||||||   ||||||         :::
Db    439 AlaSerSerIleAsnMetArgSerThrSerValSerAlaProArgThrCysArgAla--- 457

Qy   1082 ACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAA 1141
          |||   |||      :::   |||   ||||||            |||      |||
Db    458 ThrSerSerSerAlaSerCysArgCysLeuSer------------CysProGlnSerThr 473

Qy   1142 AGGCTTATAAGGAAGGAG---------CCAAGGCCA------------------------ 1168
                                     |||   |||
Db    474 ThrAlaAlaTrpAsnSerGlyTrpThrProAlaProCysProSerSerProMetAlaGly 493

Qy   1169 ACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCG---GTGGTACCGAGGTGGTTCGTT 1225
          |||   |||            :::||||||||||||         |||   |||
Db    494 ThrThrArg------------SerArgArgSerSerArgArgThrProSerTrp------ 507

Qy   1226 TTGCACAGTTTGCCAACG---CATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCA 1282
                      |||:::   :::|||                           |||
Db    508 ------------ProSerArgAsnTrp---------------------------TyrSerArg 515

Qy   1283 AATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCG 1342
          |||      |||   |||:::   :::         |||   |||   :::
Db    516 ArgArgAsnThrProSerSerAsnSerAlaLysArgArgArgThrGlyLysValSerArg 535

Qy   1343 CCATGCTCGATATTGTACGCCGGC------------------------GTATCCCTGCCG 1378
             |||:::         :::|||                        ::::::
Db    536 LysCysAlaSerThrSerSerGlyArgArgSerGlyAlaThrThrMetIleThrProThr 555

Qy   1379 ACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGA--- 1435
             |||||||||                  |||||||||   |||   |||   :::
Db    556 ValSerSerProAla---------------SerThrArgLysSerSerAlaAlaLysCys 570

Qy   1436 ---AGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCC 1492
             ||||||      |||:::         |||         |||:::      |||:::
Db    571 AlaArgSer---ProThrThrLeuValValArgSerCysArgLeuValArgLeuSerAla 589

Qy   1493 ATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGG---------CTATCGAGA 1543
                |||:::            |||      ||||||               |||
Db    590 ------CysThrTrpLysSerValArgAlaArgAlaProArgArgThrGlyAlaSerGly 607

Qy   1544 AAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCG 1603
```

```
              ||||||                                       |||   |||   |||
Db     608 ValLysArgGln---------------------------ThrThrAlaAlaPheLeuPro 618

Qy    1604 ATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGC 1663
              |||             ||||||                     |||   |||
Db     619 ---ThrTrpGluProThrArgArgGlyArgArgArgCysCysAlaArgCysProArgCys 637

Qy    1664 GTGAGA---------------TGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCA 1708
            |||               ||||||               ||||||:::
Db     638 ArgArgThrArgArgAlaThrCysThrSerAlaCysArgCysArgSer------------ 653

Qy    1709 CCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACA 1768
              |||      |||   :::   :::              ||||||      ||||||
Db     654 ---Cys------AlaProSerMetAlaSerLysSerGlyThrArgThrArgLeuGlyThr 670

Qy    1769 ACTATC-------------------------------------------ATACGACAG 1783
              |||                                        :::|||
Db     671 IleIleSerIleMetPheSerSerMetSerSerAlaArgMetLysCysAlaLeuArgSer 690

Qy    1784 GCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCG------------ 1831
            |||         ||||||   ::::::   :::   :::   |||
Db     691 AlaAsnPro------ArgSerThrMetIleLeuSerCysAlaPheProAlaThrAlaLeu 708

Qy    1832 ---------TACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATG 1882
                    :::|||   :::      |||   |||:::|||
Db     709 GlySerThrPheArgProMetGlyArgThrProSerSerProValArgAsnArgIleSer 728

Qy    1883 CCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGG------GCG 1936
                    :::||||||   :::         ||||||:::   ||||||      |||
Db     729 AlaHisProIleSerSerSerThrSerLysSerArgThrSerHisSerArgGlyAlaAla 748

Qy    1937 GTAACTCCGGTAGCCCCGTAT 1957
              |||      ||||||
Db     749 SerArgProProLeuProTyr 755
```

RESULT 4
AAB59826

```
ID   AAB59826 standard; Protein; 1615 AA.
XX
AC   AAB59826;
XX
DT   04-APR-2001  (first entry)
XX
DE   Protein #3 encoded by TutD/E gene.
XX
KW   Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 12; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ   Sequence   1615 AA;

Alignment Scores:
Pred. No.:              4.87e-07    Length:         1615
Score:                  174.00      Matches:        176
Percent Similarity:     34.00%      Conservative:   78
Best Local Similarity:  23.56%      Mismatches:     275
Query Match:            4.55%       Indels:         218
DB:                     22          Gaps:           40

US-10-008-355-1 (1-2139) x AAB59826 (1-1615)

Qy      65 CCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAA 124
            |||||||||          |||       |||         :::   ||||||:::
Db     709 ProLysProArgProThrCysArgProSerProGlyThrAlaArgArgValSerThrThr 728

Qy     125 TGCGTGAGCTCGGCTTTACGCTCCCGTTGG------------------------------ 154
                  |||            ||||||
Db     729 SerProArgSerThrThrGlyArgArgTrpSerSerProAlaArgArgSerAlaGlyArg 748

Qy     155 -----------------ATTCGCTCTACAGTTTCGACAAGCCGTCCATTG------CCA 190
                            ||||||:::      ||||||||||||:::
Db     749 AlaGlyArgAlaGlyCysAlaArgSerSerArgLysThrSerArgProIleArgSerAla 768

Qy     191 ATGCCGTGGTTATCT------------TCGGTGGCGGATGTACCG--------------- 223
              |||      |||            ||||||:::      |||
Db     769 ArgProSerCysSerLysSerProThrSerValSerAlaPheProProSerProAlaArg 788

Qy     224 GTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTA 283
              |||:::         |||      |||||||||::::::            :::
Db     789 AlaSerArgThrArgCysArgArgAsnSerLeuProSerSerValThrArgSerSerAla 808

Qy     284 TCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGA 343
            :::||||||                        |||   |||
Db     809 ThrArgAlaAlaThrProArgArgLysThrProCys------------------------ 820

Qy     344 TGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGG 403
                                                     |||   |||:::   |||
Db     821 -----------------------------------------CysGlyArgThrThrArg 826

Qy     404 TAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCA 463
                  :::      ::::::|||||||||      :::      ||||||:::
Db     827 ProProSerSerThrArgAsnSerSerArgAlaThrTrpMetArgTrpAsnSerSerArg 846

Qy     464 AAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCT 523
                 '|||:::                |||      :::       |||      :::|||
Db     847 TrpAsnValArgPheProSerMetAlaProAlaSerArgAlaProThrAlaLysSerSer 866

Qy     524 GCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCA 583
                             :::|||   :::::::::   |||::::::   |||      :::
Db     867 Arg------GlyArgThrIleCysSerSerSerProSerAlaAlaProThrProArgAla 884

Qy     584 AGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGG 643
            ||||||                           ||||||      :::      :::   :::
Db     885 ArgThr---------------------ProAlaThrThrProThrProSerSerArgGln 897

Qy     644 ACAACTGGATGTGGCCGC-----------------------GTCACACGGGCGACT--- 676
              :::|||      |||                              |||   |||
Db     898 ProSerGlySerAlaArgProSerProProSerSerSerAlaIleProArgArgThrAla 917

Qy     677 -----------------------------TCAGCGTATTCCGCGTGTATGCCGGTGCCG 706
                                        ||||||      :::   :::
Db     918 ArgArgArgCysAlaGlyPheSerSerAlaSerAlaThrAspSerAlaIleArgArgSer 937
```

```
Qy    707 ACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTG 766
          :::|||  |||        :::|||           |||  |||     |||:::
Db    938 SerThrThrArgSerAlaArgSerArgArgAsnThrProSerSerAlaSerThrAlaThr 957

Qy    767 CCGTATCCATGCAAG---------------GCTACAAGGCTGACGACTATGCCATGACCA 811
                |||   :::                 :::|||                |||
Db    958 AlaProProThrArgLysProThrThrGlySerThrCysCysAlaCysArgProAlaSer 977

Qy    812 TCGGTTTCC---------------------CGGGCAGTACGGATCGCTACCTCACTTCTT 850
          :::|||:::                     |||   |||         :::|||
Db    978 ThrValAlaAlaArgArgLysLysProValArgLysValAlaAlaGlnSerSerArgPro 997

Qy    851 GGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCA 910
                ||||||     |||   |||   |||
Db    998 SerCysTrpLysSerArgSerMetThrAlaThr--------------------------- 1008

Qy    911 AGCAAGGCATCTGGAAGGAAGCCATGAGCG---CAGATCAGGCTACCCGTA------TCA 961
                   :::||||||   |||             |||   ||||||
Db   1009 ---------ThrGlyArgThrProThrCysAsnSerAlaArgArgProValIleSerArg 1025

Qy    962 AATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACC 1021
             |||:::   |||                  ||||||:::              :::
Db   1026 ArgSerProSerArgMet---------------PheGlyArgLeu------------Ser 1036

Qy   1022 GCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAG 1081
           |||   |||:::   :::               ||||||   ||||||        :::
Db   1037 AlaSerSerIleAsnMetArgSerThrSerValSerAlaProArgThrCysArgAla--- 1055

Qy   1082 ACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAA 1141
           |||   |||       :::   |||   ||||||                |||     |||
Db   1056 ThrSerSerSerAlaSerCysArgCysLeuSer------------CysProGlnSerThr 1071

Qy   1142 AGGCTTATAAGGAAGGAG---------CCAAGGCCA------------------------ 1168
                                     |||   |||
Db   1072 ThrAlaAlaTrpAsnSerGlyTrpThrProAlaProCysProSerSerProMetAlaGly 1091

Qy   1169 ACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCG---GTGGTACCGAGGTGGTTCGTT 1225
           |||   |||              :::||||||||||||       |||   |||
Db   1092 ThrThrArg------------SerArgArgSerSerArgArgThrProSerTrp------ 1105

Qy   1226 TTGCACAGTTTGCCAACG---CATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCA 1282
                      |||:::   :::|||                        |||
Db   1106 ------------ProSerArgAsnTrp-----------------------TyrSerArg 1113

Qy   1283 AATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCG 1342
             |||   |||   |||:::   :::            |||   |||   :::
Db   1114 ArgArgAsnThrProSerSerAsnSerAlaLysArgArgArgThrGlyLysValSerArg 1133

Qy   1343 CCATGCTCGATATTGTACGCCGGC-----------------------GTATCCCTGCCG 1378
             |||:::        :::|||                           ::::::
Db   1134 LysCysAlaSerThrSerSerGlyArgArgSerGlyAlaThrThrMetIleThrProThr 1153

Qy   1379 ACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGA--- 1435
             |||||||||                 |||||||||   |||   |||   :::
Db   1154 ValSerSerProAla---------------SerThrArgLysSerSerAlaAlaLysCys 1168

Qy   1436 ---AGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCC 1492
             ||||||     |||:::         |||         |||:::     |||:::
Db   1169 AlaArgSer---ProThrThrLeuValValArgSerCysArgLeuValArgLeuSerAla 1187

Qy   1493 ATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGG---------CTATCGAGA 1543
                |||:::               |||      ||||||            |||
Db   1188 ------CysThrTrpLysSerValArgAlaArgAlaProArgArgThrGlyAlaSerGly 1205

Qy   1544 AAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCG 1603
          ||||||                               |||   |||   |||
Db   1206 ValLysArgGln--------------------------ThrThrAlaAlaPheLeuPro 1216

Qy   1604 ATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGC 1663
                |||               ||||||               |||   |||
Db   1217 ---ThrTrpGluProThrArgArgGlyArgArgArgCysCysAlaArgCysProArgCys 1235

Qy   1664 GTGAGA---------------TGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCA 1708
                  |||                 ||||||               ||||||:::
Db   1236 ArgArgThrArgArgAlaThrCysThrSerAlaCysArgCysArgSer----------- 1251

Qy   1709 CCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACA 1768
                |||      |||   :::   :::              ||||||      ||||||
Db   1252 ---Cys------AlaProSerMetAlaSerLysSerGlyThrArgThrArgLeuGlyThr 1268

Qy   1769 ACTATC------------------------------------------ATACGACAG 1783
           |||                                             :::|||
Db   1269 IleIleSerIleMetPheSerSerMetSerSerAlaArgMetLysCysAlaLeuArgSer 1288

Qy   1784 GCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCG----------- 1831
             |||         ||||||   ::::::   :::   :::   |||
Db   1289 AlaAsnPro------ArgSerThrMetIleLeuSerCysAlaPheProAlaThrAlaLeu 1306

Qy   1832 ---------TACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATG 1882
                   :::|||   :::         |||   |||:::|||
Db   1307 GlySerThrPheArgProMetGlyArgThrProSerSerProValArgAsnArgIleSer 1326

Qy   1883 CCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGG------GCG 1936
                :::||||||   :::            ||||||:::   ||||||      |||
Db   1327 AlaHisProIleSerSerSerThrSerLysSerArgThrSerHisSerArgGlyAlaAla 1346

Qy   1937 GTAACTCCGGTAGCCCCGTAT 1957
                |||      ||||||
Db   1347 SerArgProProLeuProTyr 1353
```

```
RESULT 5
AAB59827
ID   AAB59827 standard; Protein; 1592 AA.
XX
AC   AAB59827;
XX
DT   04-APR-2001  (first entry)
XX
DE   Protein #4 encoded by TutD/E gene.
XX
KW   Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 12; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
```

```
XX
SQ   Sequence   1592 AA;

Alignment Scores:
Pred. No.:              3.74e-05        Length:         1592
Score:                  154.50          Matches:        140
Percent Similarity:     28.05%          Conservative:   46
Best Local Similarity:  21.12%          Mismatches:     172
Query Match:            3.90%           Indels:         305
DB:                     22              Gaps:           38

US-10-008-355-1 (1-2139) x AAB59827 (1-1592)

Qy     2080 AGAGAACGTAGCGGATGTCCACGCTGATTGTGCGCTGCAGATCGGGTTCGAACTCGATGT 2021
            |||   |||      |||   |||                        |||   ||||||
Db      768 ArgCysArgAspAlaCysGluArg------------------------ArgAlaArgCys 779

Qy     2020 CACCACTCATAGCTTCCCAGTTGCCATCGAAAGCAAGACCGATCAGACGGCCGTTCTTAT 1961
                                       |||:::      |||          |||:::
Db      780 ProGlyProArgSerAlaProSerIleArgArgGlySerArgAspArgSerArgAlaSer 799

Qy     1960 CGAATACGGGGCTACCGGAGTTACCGCCCCGTGATGT------------------------ 1925
            |||   ||||||               |||   |||
Db      800 ArgSerArgGly------------SerProLeuCysGlyAlaThrAlaThrSerCysPro 815

Qy     1924 ---------CGTTGTTCGATAGGA------------------------------AAGCGA 1904
                     ||||||||||||||                               :::|||
Db      816 ArgArgArgArgCysSerIleGlyAlaSerSerGlyCysProHisProProValArgArg 835

Qy     1903 TATGGAGCTGACCGTTCTCGGCATAGCGACCATAGTTTTGGTGCGGAAGAGGTCGAGGA 1844
                     :::   |||:::   :::                     |||||||||
Db      836 SerProValAsnSerSerLysArgAlaHisArgArgCysThrAlaArgArgGlyArgPhe 855

Qy     1843 TATTCTCCTGTA---CGGCAAACTCATCGCTCTTAGGATCCTGCTTCTCCAATACGCCCT 1787
                  |||      |||   |||   |||                        |||
Db      856 ArgGlyProThrSerArgAspThrGlyArg------------------------ArgArg 867

Qy     1786 TGC------------------------------------------------------CTG 1781
            |||                                                      |||
Db      868 CysTrpArgTrpProArgProArgArgCysArgCysSerArgArgTrpGlyArgProLeu 887

Qy     1780 TCGTATGATAGTTGTACCAGGCACCGTCCTGCGGTTCATATCCCTTGATGGAGCCGTAGC 1721
                        |||   |||   |||                     |||   |||
Db      888 TrpAlaSerGlyCysProArgAlaArg------------------------TrpArgGly 900

Qy     1720 TCATACGCATGGTGAAGTTGGCATCGCTCGGCA------------------------GAG 1685
            |||      |||   |||         ||||||
Db      901 Ser---AsnTrpSerSerGlyArgSerSerAlaAlaSerProLysArgThrCysGlyArg 919

Qy     1684 CACGTCCGGGGTACATCTCACGCAAACCGGCAAAGAAAAGACGCTTGCCCTTCTCAATGG 1625
               ||||||      :::   |||   |||   |||         ||||||:::
Db      920 ArgValArgSerAspThrSerAlaArgArgSerArg---------CysProAla------ 934

Qy     1624 CATAGGCATTGGCCATCGCATCGGCCTGAATAGCGCGAGCAGCAGCTATTACGCTCTTGG 1565

Db      934 ------------------------------------------------------------ 934

Qy     1564 AAAGCTCTACTGCCGGATCTTTCTCGATAGCCTTGGCAAACTTTTCCTTGTCCATGGACT 1505
                                                   :::   |||      ||||||
Db      935 ---------------------------------------SerSerProIleArgTrpThr 941

Qy     1504 TGAGCATGGCATGGAACTTGT---CGCTATAAGGAACCACACTCTTGTCGAATACGAAGT 1448
                        |||   |||   |||::::::   |||      |||      |||:::
Db      942 ------------GlyArgCysArgArgTrpArgArgProLeuGlyCysSerProArgAla 957

Qy     1447 CTGCATACTTCTTCGTGTCGCCTTTGAATTTCTTGTCGATTACATTCTTGAATATATCGG 1388
                  |||:::   |||                                       |||
Db      958 ThrCysThrAlaArgCysGly------------------------------------Arg 965

Qy     1387 GGAGCTTGTCGGCAGGGATACGCCGGCGTACAATATCGAGCATGGCGG---------GCA 1337
                  |||         :::            :::|||:::   |||         |||
Db      966 AspGlyCysSerAlaPhePheGlyAsnProLeuHisArgSerLeuArgGlyProTrpAla 985

Qy     1336 GCACCTTACGGTCGAGCGAGGGGAGGTAGTCTTTGTACTTGTCGTCAAGCGATTTGAGGA 1277
            ||||||:::   ||||||                  |||   |||   |||:::   |||
Db      986 AlaProPhe---ArgAlaHisArgSerArgSerThrThrArgArgCysAlaValArgGly 1004

Qy     1276 ---TACCGGCATGAGCAT------------------------------------------ 1262
                  ||||||:::
Db     1005 SerSerArgHisAspArgThrAlaSerThrArgArgProHisLysProProLysGlyCys 1024

Qy     1261 ---CAGGATTTGTAGCCAATGCGTTGG---------CAAACTGTGCAAAACGAA------ 1220
                  |||:::         |||:::         :::|||      :::
Db     1025 AlaThrAspIleHisSerGlyArgTyrCysTrpProArgThrAlaSerSerArgAlaAla 1044

Qy     1219 ---------------------------------------------CCACCTCGGTACCAC 1205
                                                         |||   |||
Db     1045 SerGlyAlaSerAlaLysArgThrArgLeuArgArgArgSerCysProValArgSerPro 1064

Qy     1204 CGAAGA---------------------------------GCGTCTCGCTCAAATAAGTCA 1178
            ||||||                                    |||   |||   :::
Db     1065 ArgArgArgGlyThrArgAlaAlaTrpHisSerAlaCysGlySerSerSerArgArgPro 1084

Qy     1177 TCTCACGGTTGGCCTTGGCTCCTTCCTTAT------------------------------ 1148
            |||   |||   ||||||   :::|||
Db     1085 SerSerGlyArgProTrpSerValProIleArgProSerSerIleCysGlyArgAlaVal 1104

Qy     1147 ------------------------------AAGCCTTTTCGAGAGAAGACAATACATCGC 1118
                                          :::||||||      ::::::
Db     1105 GlyLeuThrSerProSerSerProLeuAsnArgProPheAlaArgArgSerAlaProAla 1124

Qy     1117 CAT---------AGACAGCACTCTTGCCGTTCTTACGGATCCAGTCTGCGAATGCTCTTT 1067
                        |||:::|||:::   |||   |||||||||      |||      |||
Db     1125 SerThrProCysArgArgHisAsnArgArgArgTyrGlySer------ArgArgProPhe 1142

Qy     1066 CCTCGGCAC---------------------GCTTACGACCTATCACGTCAAGACGAGCGA 1028
               |||                           :::|||   :::   ||||||   :::
Db     1143 ArgArgArgPheAlaCysSerTrpSerSerGlnHisAspProAlaSerGlnAspProGln 1162

Qy     1027 GACCGCGGTTCATACCGATCGAATTCTTCCAATAGTTAGCACTCTGAGCATACTTGCTGG 968
            ||||||                                             ||||||
Db     1163 ---ArgGly------------------------------------------ThrCys--- 1166

Qy      967 CATATTTGATACGGGTAGCCTGATCTGCGCTCA------------TGGCTTCCTTCCAGA 920
                                    ||||||:::            |||   |||
Db     1167 ---------------------ProLeuArgAsnAlaCysProGlyTrpAlaPro------ 1177

Qy      919 TGCCTTGCTTGATACCGCGAACTTCGATACGAGGATTGTTCTCGTTTTCGATACGATCTT 860
                                                            |||
Db     1178 ------------------------------------------------ArgAlaSerArg 1181

Qy      859 CCACAC---CCCAAGAAGTGAGGTAGCGATCCGTACTGCCCG------------------ 821
            ||||||   |||   :::                  ||||||
Db     1182 ProHisLeuProLeuArgArgArgCysLysGluArgCysProPheArgCysSerProAla 1201

Qy      820 ---------------GGAAACCGATGGTCATGGCATAGTCGTCAGCCTTGT--------- 785
                                    |||         |||      ||||||
Db     1202 SerProProThrAlaSerProThrTrpProAlaSerSerGlySerProCysGlyAlaSer 1221

Qy      784 ------------------AGCCTTGCATGGATACGGCAGCGA------------------ 761
                              |||   :::|||         |||
Db     1222 GlyAlaSerThrProAlaSerAlaSerTrpAlaHisSerArgPheArgSerSerThrCys 1241

Qy      760 ---------AGTAAACGGGCTTATAGGGTTTAT------------TGTCCT--------- 731
                     ::::::   :::||||||   :::            ||||||
Db     1242 SerSerLysAlaArgAspSerTyrArgAspHisGluThrHisArgCysProHisPheAla 1261

Qy      730 ---------TGCTGTATTCGGCCGGCCGGTTGTCGGCACCGGCATACACGCGGAATACGC 680
                           :::|||   ||||||                  |||      |||
Db     1262 ArgGlnAlaAlaArgValArgSerAlaGlyHisAspArgAlaArgLysArgSerCysArg 1281

Qy      679 TGAAGTCGCCCGTGTGACGCGGCCACATCCAGTTGT------------------CCGTAT 638
            :::|||   |||::::::         :::||||||                  |||
Db     1282 GluThrArgLeuCysGluSerArgHisAlaSerCysTrpArgGlyAspProValProAsn 1301

Qy      637 CGCCTCCGA 629
```

```
           |||:::|||
Db    1302 ArgMetArg 1304

RESULT 6
AAB59817
ID   AAB59817 standard; Protein; 999 AA.
XX
AC   AAB59817;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #8.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;    99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   999 AA;

Alignment Scores:
Pred. No.:              3.37e-05      Length:        999
Score:                  153.50        Matches:       143
Percent Similarity:     28.27%        Conservative:  45
Best Local Similarity:  21.50%        Mismatches:    170
Query Match:            3.88%         Indels:        307
DB:                     22            Gaps:          40

US-10-008-355-1 (1-2139) x AAB59817 (1-999)

Qy    2080 AGAGAACGTAGCGGATGTCCACGCTGATTGTGCGCTGCAGATCGGGTTCGAACTCGATGT 2021
           |||  |||           |||  |||                     |||  ||||||
Db     173 ArgCysArgAspAlaCysGluArg------------------------ArgAlaArgCys 184

Qy    2020 ---CACCACTCATAGCTTCCCAGTTGCCATCGA---AAGCAAGACCGATCAGACGGCCGT 1967
                             ||||||      |||      :::|||||||||
Db     185 ProGlyProArgSerAlaProSerIleArgArgGlySerArgAspArgSerArgAlaSer 204

Qy    1966 TCTTATCGAATACGGGGCTACCGGAGTTACCGCCCGTGATGT------------------ 1925
              |||  ||||||                        |||   |||
Db     205 ArgSerArgSerArgGly------------SerProLeuCysGlyAlaThrAlaThrSer 220

Qy    1924 ---------------CGTTGTTCGATAGGA------------------------------ 1910
                          ||||||||||||||
Db     221 CysProArgArgArgArgCysSerIleGlyAlaSerSerGlyCysProHisProProVal 240

Qy    1909 AAGCGATATGGAGCTGACCGTTCTCGGCATAGCGACCATAGTTTTTGGTGCGGAAGAGGT 1850
           :::|||       :::    |||:::    :::                     ||||||
Db     241 ArgArgSerProValAsnSerSerLysArgAlaHisArgArgCysThrAlaArgArgGly 260

Qy    1849 CGAGGATATTCTCCTGTA---CGGCAAACTCATCGCTCTTAGGATCCTGCTTCTCCAATA 1793
           |||       |||        |||   |||   |||
Db     261 ArgPheArgGlyProThrSerArgAspThrGlyArg------------------------ 272

Qy    1792 CGCCCTTGC--------------------------------------------------- 1784
           |||   |||
Db     273 ArgArgCysTrpArgTrpProArgProArgArgCysArgCysSerArgArgTrpGlyArg 292

Qy    1783 ---CTGTCGTATGATAGTTGTACCAGGCACCGTCCTGCGGTTCATATCCCTTGATGGAGC 1727
              |||            |||   |||   |||                     |||
Db     293 ProLeuTrpAlaSerGlyCysProArgAlaArg---------------------TrpArg 305

Qy    1726 CGTAGCTCATACGCATGGTGAAGTTGGCATCGCTCGGCA--------------------- 1688
           |||   |||      |||   |||         ||||||
Db     306 ArgGlySer---AsnTrpSerSerGlyArgSerSerAlaAlaSerProLysArgThrCys 324

Qy    1687 ---GAGCACGTCCGGGGTACATCTCACGCAAACCGGCAAAGAAAAGACGCTTGCCCTTCT 1631
                   ||||||      :::   |||   |||   |||         ||||||:::
Db     325 GlyArgArgValArgSerAspThrSerAlaArgArgSerArg---------CysProAla 341

Qy    1630 CAATGGCATAGGCATTGGCCATCGCATCGGCCTGAATAGCGCGAGCAGCAGCTATTACGC 1571

Db     341 ------------------------------------------------------------ 341

Qy    1570 TCTTGGAAAGCTCTACTGCCGGATCTTTCTCGATAGCCTTGGCAAACTTTTCCTTGTCCA 1511
                                                         :::   |||
Db     342 ---------------------------------------------SerSerProIleArg 346

Qy    1510 TGGACTTGAGCATGGCATGGAACTTGT---CGCTATAAGGAACCACACTCTTGTCGAATA 1454
           ||||||            |||   |||   |||::::::   |||      |||
Db     347 TrpThr------------GlyArgCysArgArgTrpArgArgProLeuGlyCysSerPro 362

Qy    1453 CGAAGTCTGCATACTTCTTCGTGTCGCCTTTGAATTTCTTGTCGATTACATTCTTGAATA 1394
           |||:::      |||:::   |||
Db     363 ArgAlaThrCysThrAlaArgCysGly--------------------------------- 371

Qy    1393 TATCGGGGAGCTTGTCGGCAGGGATACGCCGGCGTACAATATCGAGCATGGCGG------ 1340
              |||      |||         :::         :::|||:::   |||
Db     372 ---ArgAspGlyCysSerAlaPhePheGlyAsnProLeuHisArgSerLeuArgGlyPro 390

Qy    1339 ---GCAGCACCTTACGGTCGAGCGAGGGGAGGTAGTCTTTGTACTTGTCGTCAAGCGATT 1283
              |||||||||:::   ||||||               |||   |||   |||:::
Db     391 TrpAlaAlaPhe---ArgAlaHisArgSerArgSerThrThrArgArgCysAlaVal 409

Qy    1282 TGAGGA---TACCGGCATGAGCAT------------------------------------ 1262
              |||      ||||||:::
Db     410 ArgGlySerSerArgHisAspArgThrAlaSerThrArgArgProHisLysProProLys 429

Qy    1261 ---------CAGGATTTGTAGCCAATGCGTTGG---------CAAACTGTGCAAAACGAA 1220
                    |||:::         |||:::         :::|||      :::
Db     430 GlyCysAlaThrAspIleHisSerGlyArgTyrCysTrpProArgThrAlaSerSerArg 449

Qy    1219 ---------------------------------------------------CCACCTCGG 1211
                                                              |||   |||
Db     450 AlaAlaSerGlyAlaSerAlaLysArgThrArgLeuArgArgArgSerCysProValArg 469

Qy    1210 TACCACCGAAGA------------------------------------GCGTCTCGCTCAAAT 1184
              ||||||                                   |||   |||
Db     470 SerProArgArgArgGlyThrArgAlaAlaTrpHisSerAlaCysGlySerSerSerArg 489

Qy    1183 AAGTCATCTCACGGTTGGCCTTGGCTCCTTCCTTAT------------------------ 1148
           :::   |||   |||   ||||||   :::|||
Db     490 ArgProSerSerGlyArgProTrpSerValProIleArgProSerSerIleCysGlyArg 509
```

```
Qy   1147 -----------------------------------AAGCCTTTTCGAGAGAAGACAATA 1124
                                             :::||||||     :::::: 
Db    510 AlaValGlyLeuThrSerProSerSerProLeuAsnArgProPheAlaArgArgSerAla 529

Qy   1123 CATCGCCAT---------AGACAGCACTCTTGCCGTTCTTACGGATCCAGTCTGCGAATG 1073
                   |||:::|||:::   |||   |||||||||      |||
Db    530 ProAlaSerThrProCysArgArgHisAsnArgArgArgTyrGlySer------ArgArg 547

Qy   1072 CTCTTTCCTCGGCAC-------------------GCTTACGACCTATCACGTCAAGAC 1034
             |||   |||                     :::|||   :::   ||||||
Db    548 ProPheArgArgArgPheAlaCysSerTrpSerSerGlnHisAspProAlaSerGlnAsp 567

Qy   1033 GAGCGAGACCGCGGTTCATACCGATCGAATTCTTCCAATAGTTAGCACTCTGAGCATACT 974
             :::   ||||||                                        |||
Db    568 ProGln---ArgGly---------------------------------------Thr 572

Qy    973 TGCTGGCATATTTGATACGGGTAGCCTGATCTGCGCTCA-----------TGGCTTCCT 926
          |||                        ||||||:::           |||   |||
Db    573 Cys------------------------ProLeuArgAsnAlaCysProGlyTrpAlaPro 584

Qy    925 TCCAGATGCCTTGCTTGATACCGCGAACTTCGATACGAGGATTGTTCTCGTTTTCGATAC 866
                                                                  |||
Db    585 --------------------------------------------------------ArgAla 586

Qy    865 GATCTTCCACAC---CCCAAGAAGTGAGGTAGCGATCCGTACTGCCCG----------- 821
                ||||||   |||   :::                   ||||||
Db    587 SerArgProHisLeuProLeuArgArgArgCysLysGluArgCysProPheArgCysSer 606

Qy    820 --------------------GGAAACCGATGGTCATGGCATAGTCGTCAGCCTTGT--- 785
                                 |||         |||      ||||||
Db    607 ProAlaSerProProThrAlaSerProThrTrpProAlaSerSerGlySerProCysGly 626

Qy    784 -----------------------AGCCTTGCATGGATACGGCAGCGA----------- 761
                                 |||   :::|||         |||
Db    627 AlaSerGlyAlaSerThrProAlaSerAlaSerTrpAlaHisSerArgPheArgSerSer 646

Qy    760 --------------AGTAAACGGGCTTATAGGGTTTAT-----------TGTCCT--- 731
                        ::::::   :::||||||   :::            ||||||
Db    647 ThrCysSerSerLysAlaArgAspSerTyrArgAspHisGluThrHisArgCysProHis 666

Qy    730 --------------TGCTGTATTCGGCCGGCCGGTTGTCGGCACCGGCATACACGCGGA 686
                                 :::|||   ||||||                |||
Db    667 PheAlaArgGlnAlaAlaArgValArgSerAlaGlyHisAspArgAlaArgLysArgSer 686

Qy    685 ATACGCTGAAGTCGCCCGTGTGACGCGGCCACATCCAGTTGT----------------- 644
             |||   :::|||   |||::::::      :::||||||
Db    687 CysArgGluThrArgLeuCysGluSerArgHisAlaSerCysTrpArgGlyAspProVal 706

Qy    643 CCGTATCGCCTCCGA 629
          |||   |||:::|||
Db    707 ProAsnArgMetArg 711
```

RESULT 7
ABG66702

```
ID   ABG66702 standard; Protein; 1296 AA.
XX
AC   ABG66702;
XX
DT   30-AUG-2002  (first entry)
XX
DE   Human novel polypeptide #37.
XX
KW   Human; inflammatory condition; shock; sepsis; immune response;
KW   cancer; wound healing; central nervous system disease; haematopoiesis;
KW   peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW   myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW   cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW   bone degenerative disorder; periodontal disease; reperfusion injury;
KW   lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW   allergic condition; thrombolysis; thrombosis; coagulation disorder;
KW   fungal infection.
XX
OS   Homo sapiens.
XX
PN   WO200244340-A2.
XX
PD   06-JUN-2002.
XX
PF   30-NOV-2001; 2001WO-US47004.
XX
PR   30-NOV-2000; 2000US-0028952.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Goodrich RW,  Liu C,  Zhou P,  Asundi V,  Wang J,  Wang D;
PI   Yamazaki V,  Ujwal ML,  Drmanac RT;
XX
DR   WPI; 2002-508509/54.
DR   N-PSDB; ABK94926.
XX
PT   Novel nucleic acids and polypeptides for diagnosis, treatment of
PT   inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT   disorders, cancer and promoting wound healing -
XX
PS   Claim 10; Page 604-607; 672pp; English.
XX
CC   The invention relates to human novel polynucleotides and associated
CC   polypeptides. The polynucleotides and polypeptides are useful for
CC   treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC   and cancer and for promoting wound healing. The sequences are used to
CC   induce the proliferation of neural cells and regeneration of nerve and
CC   brain tissue, and are useful for the treatment of central and peripheral
CC   nervous system diseases and neuropathies, such as Alzheimer's disease,
CC   Parkinson's disease, Huntington's disease and amyotrophic lateral
CC   sclerosis. The sequences are involved in chemotactic or chemokinetic
CC   activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC   cell disorders and platelet disorders such as thrombocytopenia,
CC   regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC   growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC   osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC   disease. The sequences of the invention are also useful for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues, immune deficiencies and disorders
CC   including severe combined immunodeficiency (SCID), bacterial or fungal
CC   infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC   gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC   and coagulation disorders. Sequences ABG66666-ABG66758 represent human
CC   novel polypeptides of the invention.
XX
SQ   Sequence   1296 AA;
```

```
Alignment Scores:
Pred. No.:              4.05e-05      Length:        1296
Score:                  153.50        Matches:       150
Percent Similarity:     33.76%        Conservative:  86
Best Local Similarity:  21.46%        Mismatches:    251
Query Match:            4.02%         Indels:        212
DB:                     23            Gaps:          27
```

US-10-008-355-1 (1-2139) x ABG66702 (1-1296)

```
Qy    506 CAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCA 565
          :::|||         |||      :::   |||   ||||||   |||::::::
Db    539 LysThrThrLeuThrSerLeuLysThrThrAlaSerArgProThrAlaAsnSerThrLeu 558

Qy    566 TCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTA 625
          ||||||      |||                     :::|||   :::      |||
Db    559 SerSerLeuThrSerSer---------------------IleLeuSerSerThrLeuVal 571

Qy    626 AGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTAT 685
             |||      :::   |||:::            :::||||||   :::|||
Db    572 ProSerThrAspMetIleThrSerHisThrThrAsnLeuThrArg---SerSerProLeu 590

Qy    686 TCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCT 745
             |||   :::|||         |||         |||:::   |||      |||
```

```
Db     591 LeuAlaThrLeuProThrThrIleThrArgSerThrProThrSerGluThrThrTyrPro 610

Qy     746 ATAAGCCCG--------------------TTTACTTCGCTGCCGTATCCATGCAAGGCT 784
           ||||||                        ||||||:::   |||
Db     611 ThrSerProThrSerThrValLysGlySerThrThrSerIleArgTyrSerThrSerMet 630

Qy     785 ACAAGGCTG------------ACGACTATGCCA---------------TGACCATCGGTT 817
           |||                 |||:::::|||                |||:::
Db     631 ThrGlyThrLeuSerMetGluThrSerLeuProProThrSerSerSerLeuProThrThr 650

Qy     818 TCCCGGGCAGTACGGATCGCTACCTCACTTCTT-------------------------- 850
              |||             |||:::  |||
Db     651 GluThrAlaThrMetThrProThrThrThrLeuIleThrThrThrProAsnThrThrSer 670

Qy     851 -------------------------GGGGTGTGGAAGATCGTATCGAAAACGAGAACA 883
                                     ::::::   ||||||  :::  |||
Db     671 HisSerThrProSerPheThrSerSerThrIleTyrSerThrValSerThrSerThrThr 690

Qy     884 ATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAG 943
              :::  |||  |||      |||     |||          |||
Db     691 AlaIleThrSerHisPheThrThrSerGluThrAlaValThrProThrProValThrPro 710

Qy     944 ATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGA 1003
              |||         :::::::::|||   :::  ||||||   ||||||
Db     711 SerSerLeuSer---ThrAspIleProThrThrSerLeuArgThrLeuThrProSer--- 728

Qy    1004 ATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTG--- 1060
              ||||||  |||:::  ||||||       |||        |||:::
Db     729 ------SerValGlyThrSerThrSerLeuThrThrThrThrAspPheProSerIlePro 746

Qy    1061 -----------------CCGAGGAAAGAGCAT------TCGCAGACTGGATCCGTAAGA 1096
                            |||   :::   |||      :::   |||::::::
Db     747 ThrAspIleSerThrLeuProThrArgThrHisIleIleSerSerSerProSerIleGln 766

Qy    1097 ACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAG 1156
           :::            |||:::                :::        :::|||
Db     767 SerThrGluThrSerSerLeuValGlyThrThrSerProThrMetSerThrValArgMet 786

Qy    1157 GAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGG 1216
                 |||   |||           |||   :::      :::|||:::|||
Db     787 ThrLeuArgIleThrGluAsnThrProIleSerSerPheSerThrSerIleVal------ 804

Qy    1217 TGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTA 1276
                 |||:::          ||||||                         ||||||
Db     805 ------ValIleProGluThrProThr------------------GlnThrProProVal 816

Qy    1277 TCCTCAAATCGCTTGACGACAAGTACA-----------------------------AAG 1306
             ::::::         |||  |||
Db     817 LeuThrSerAlaThrGlyThrGlnThrSerProAlaProThrThrValThrPheGlySer 836

Qy    1307 ACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGC 1366
           |||   |||     ||||||:::                |||   :::   |||
Db     837 ThrAspSerSerThrSerThrLeuHisThrLeuThrProSerThrAlaLeuSerThrIle 856

Qy    1367 GTATCC---------------CTGCCG------------------------------- 1378
           ||||||                :::|||
Db     857 ValSerThrSerGlnValProIleProSerThrHisSerSerThrLeuGlnThrThrPro 876

Qy    1379 -------------------------ACAAGCTCCCCGATATATTCAAGAATGTAATCG 1411
                                    ||||||:::        ::::::
Db     877 SerThrProSerLeuGlnThrSerLeuThrSerThrSerGluPheThrThrGluSerPhe 896

Qy    1412 ACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGG 1471
           ||||||  |||   :::|||      ::::::   |||||||:::||||||   :::|||
Db     897 ThrArgGlySerThrSerThr---AsnAlaIleLeuThrSerPheSerThrIleIleTrp 915

Qy    1472 TTC---------CTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAA 1522
                          |||    ||||||   |||   |||:::       |||
Db     916 SerSerThrProThrIleIleMetSerSerSerProSerSerAlaSerIleThr--ProV 935

Qy    1523 AGTTTGCCAAGGCTATCGAGAAAGATCCGGCA-------------------------- 1554
             |||:::      |||         |||:::
Db     935 alPheSerThrThrIleHisSerValProSerSerProTyrIlePheSerThrGluAsnV 955

Qy    1555 ------------------------GTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTC 1588
                                   ||||||   |||          :::
Db     955 alGlySerAlaSerIleThrGlyPheProSerLeuSerSerSerAlaThrThrSerThrS 975

Qy    1589 GCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTT 1648
              :::         :::                            |||         |||
Db     975 erSerThrSerSerSerLeuThrThrAlaLeuThrGluIleThrPro-PheSerTyrIle 994

Qy    1649 TCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTG---------CTCTGCCGAGCGATG 1699
           |||||||||            |||          :::         :::        |||
Db     995 SerLeuProSerThrThrProCysProGlyThrIleThrIleThrIleValProAlaSer 1014

Qy    1700 CCAACTTCACCATGCGTA-----------------TGAGCTACGGCTCCA--------- 1732
           ||||||   |||||||||                ||||||:::|||
Db    1015 ProThrAspProCysValGluMetAspProSerThrGluAlaThrSerProProThrThr 1034

Qy    1732 ------------------------------------------------------------ 1732

Db    1035 ProLeuThrValPheProPheThrThrGluMetValThrCysProThrSerIleSerIle 1054

Qy    1732 ------------------------------------------------------------ 1732

Db    1055 GlnThrThrLeuThrThrTyrMetAspThrSerSerMetMetProGluSerGluSerSer 1074

Qy    1733 --------------TCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATA 1777
                         |||                |||||||||      |||:::
Db    1075 IleSerProAsnAlaSerSerSerThrGlyThrGlyThrValProThrAsnThrValPhe 1094

Qy    1778 CGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGG 1837
                 |||                                    ||||||
Db    1095 ThrSerThrArg-----------------------------------LeuPro------ 1100

Qy    1838 AGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACG---GTC 1894
              |||   |||      :::        :::::::::   :::|||
Db    1101 ---ThrSerGluThrTrpLeuSerAsnSerSerValIleProLeuProLeuProGlyVal 1119

Qy    1895 AGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCG 1954
           |||:::|||   |||:::   :::   ::::::|||                  |||
Db    1120 SerThrIleProLeuThrMetLysProSerSerSer-----------------LeuPro 1133

Qy    1955 TATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGA 2014
                 :::|||||||:::      |||          |||       |||    :::
Db    1134 ThrIleLeuArgThrSerSerLysSerThrHisProSerProProThrThrArg------ 1151

Qy    2015 GTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACG 2074
                    :::|||     |||:::   :::   |||          |||      |||
Db    1152 ---------ThrSerGluThrProValAlaThrThrGln---------ThrProThrThr 1165

Qy    2075 TTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
              :::|||     |||     :::::::::   ::::::::::|||
Db    1166 LeuThrSerArgArgThrThrArgIleThrSerGlnMetThrThrGlnSer 1182
```

```
RESULT 8
ABG66756
ID   ABG66756 standard; Protein; 1296 AA.
XX
AC   ABG66756;
XX
DT   30-AUG-2002  (first entry)
XX
DE   Human novel polypeptide #91.
XX
KW   Human; inflammatory condition; shock; sepsis; immune response;
KW   cancer; wound healing; central nervous system disease; haematopoiesis;
KW   peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW   myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW   cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW   bone degenerative disorder; periodontal disease; reperfusion injury;
KW   lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW   allergic condition; thrombolysis; thrombosis; coagulation disorder;
KW   fungal infection.
XX
```

```
OS   Homo sapiens.
XX
PN   WO200244340-A2.
XX
PD   06-JUN-2002.
XX
PF   30-NOV-2001; 2001WO-US47004.
XX
PR   30-NOV-2000; 2000US-0028952.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Goodrich RW,  Liu C,  Zhou P,  Asundi V,  Wang J,  Wang D;
PI   Yamazaki V,  Ujwal ML,  Drmanac RT;
XX
DR   WPI; 2002-508509/54.
DR   N-PSDB; ABK94980.
XX
PT   Novel nucleic acids and polypeptides for diagnosis, treatment of
PT   inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT   disorders, cancer and promoting wound healing -
XX
PS   Claim 10; Page 669-671; 672pp; English.
XX
CC   The invention relates to human novel polynucleotides and associated
CC   polypeptides. The polynucleotides and polypeptides are useful for
CC   treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC   and cancer and for promoting wound healing. The sequences are used to
CC   induce the proliferation of neural cells and regeneration of nerve and
CC   brain tissue, and are useful for the treatment of central and peripheral
CC   nervous system diseases and neuropathies, such as Alzheimer's disease,
CC   Parkinson's disease, Huntington's disease and amyotrophic lateral
CC   sclerosis. The sequences are involved in chemotactic or chemokinetic
CC   activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC   cell disorders and platelet disorders such as thrombocytopenia,
CC   regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC   growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC   osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC   disease. The sequences of the invention are also useful for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues, immune deficiencies and disorders
CC   including severe combined immunodeficiency (SCID), bacterial or fungal
CC   infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC   gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC   and coagulation disorders. Sequences ABG66666-ABG66758 represent human
CC   novel polypeptides of the invention.
XX
SQ   Sequence   1296 AA;

Alignment Scores:
Pred. No.:              4.05e-05      Length:         1296
Score:                  153.50        Matches:        150
Percent Similarity:     33.76%        Conservative:   86
Best Local Similarity:  21.46%        Mismatches:     251
Query Match:            4.02%         Indels:         212
DB:                     23            Gaps:           27

US-10-008-355-1 (1-2139) x ABG66756 (1-1296)

Qy     506 CAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCA 565
               :::|||      |||       :::   |||  ||||||  |||:::::::
Db     539 LysThrThrLeuThrSerLeuLysThrThrAlaSerArgProThrAlaAsnSerThrLeu 558

Qy     566 TCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTA 625
           ||||||    |||                    :::|||   :::      |||
Db     559 SerSerLeuThrSerSer--------------------IleLeuSerSerThrLeuVal 571

Qy     626 AGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTAT 685
              |||   :::   |||:::            :::||||||   :::|||
Db     572 ProSerThrAspMetIleThrSerHisThrThrAsnLeuThrArg---SerSerProLeu 590

Qy     686 TCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCT 745
              |||    :::|||         |||             |||:::   |||     |||
Db     591 LeuAlaThrLeuProThrThrIleThrArgSerThrProThrSerGluThrThrTyrPro 610

Qy     746 ATAAGCCCG--------------------TTTACTTCGCTGCCGTATCCATGCAAGGCT 784
              ||||||                         ||||||:::   |||
Db     611 ThrSerProThrSerThrValLysGlySerThrThrSerIleArgTyrSerThrSerMet 630

Qy     785 ACAAGGCTG-----------ACGACTATGCCA---------------TGACCATCGGTT 817
           |||               |||:::::|||                     |||:::
Db     631 ThrGlyThrLeuSerMetGluThrSerLeuProProThrSerSerSerLeuProThrThr 650

Qy     818 TCCCGGGCAGTACGGATCGCTACCTCACTTCTT-------------------------- 850
                 |||             |||:::   |||
Db     651 GluThrAlaThrMetThrProThrThrThrLeuIleThrThrThrProAsnThrThrSer 670

Qy     851 --------------------------GGGGTGTGGAAGATCGTATCGAAAACGAGAACA 883
                                        ::::::     ||||||   :::   |||
Db     671 HisSerThrProSerPheThrSerSerThrIleTyrSerThrValSerThrSerThrThr 690

Qy     884 ATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAG 943
              :::   |||   |||         |||         |||            |||
Db     691 AlaIleThrSerHisPheThrThrSerGluThrAlaValThrProThrProValThrPro 710

Qy     944 ATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGA 1003
                 |||         ::::::::|||    :::   ||||||   ||||||
Db     711 SerSerLeuSer---ThrAspIleProThrThrSerLeuArgThrLeuThrProSer--- 728

Qy    1004 ATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTG--- 1060
                 ||||||   |||:::   ||||||      |||             |||:::
Db     729 ------SerValGlyThrSerThrSerLeuThrThrThrThrAspPheProSerIlePro 746

Qy    1061 -----------------CCGAGGAAAGAGCAT------TCGCAGACTGGATCCGTAAGA 1096
                             |||   :::   |||     |||   :::   |||::::::
Db     747 ThrAspIleSerThrLeuProThrArgThrHisIleIleSerSerSerProSerIleGln 766

Qy    1097 ACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAG 1156
              :::            |||:::                  :::         :::|||
Db     767 SerThrGluThrSerSerLeuValGlyThrThrSerProThrMetSerThrValArgMet 786

Qy    1157 GAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGG 1216
                 |||   |||            |||   :::      :::|||:::|||
Db     787 ThrLeuArgIleThrGluAsnThrProIleSerSerPheSerThrSerIleVal------ 804

Qy    1217 TGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTA 1276
                 |||:::         ||||||                           ||||||
Db     805 ------ValIleProGluThrProThr-----------------GlnThrProProVal 816

Qy    1277 TCCTCAAATCGCTTGACGACAAGTACA-----------------------------AAG 1306
              ::::::         |||   |||
Db     817 LeuThrSerAlaThrGlyThrGlnThrSerProAlaProThrThrValThrPheGlySer 836

Qy    1307 ACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGC 1366
           |||   |||      |||||||:::                  |||   :::   |||
Db     837 ThrAspSerSerThrSerThrLeuHisThrLeuThrProSerThrAlaLeuSerThrIle 856

Qy    1367 GTATCC---------------CTGCCG-------------------------------- 1378
           ||||||               :::|||
Db     857 ValSerThrSerGlnValProIleProSerThrHisSerSerThrLeuGlnThrThrPro 876

Qy    1379 --------------------------ACAAGCTCCCCGATATATTCAAGAATGTAATCG 1411
                                     ||||||:::       ::::::
Db     877 SerThrProSerLeuGlnThrSerLeuThrSerThrSerGluPheThrThrGluSerPhe 896

Qy    1412 ACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGG 1471
           ||||||   |||   :::|||      ::::::   ||||||:::||||||    :::|||
Db     897 ThrArgGlySerThrSerThr---AsnAlaIleLeuThrSerPheSerThrIleIleTrp 915

Qy    1472 TTC---------CTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAA 1522
                       |||      ||||||   |||   |||:::      |||
Db     916 SerSerThrProThrIleIleMetSerSerSerProSerSerAlaSerIleThr--ProV 935

Qy    1523 AGTTTGCCAAGGCTATCGAGAAAGATCCGGCA-------------------------- 1554
              |||:::         |||         |||:::
```

```
Db      935 alPheSerThrThrIleHisSerValProSerSerProTyrIlePheSerThrGluAsnV 955

Qy     1555 -------------------------GTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTC 1588
                                         ||||||  |||          :::
Db      955 alGlySerAlaSerIleThrGlyPheProSerLeuSerSerSerAlaThrThrSerThrS 975

Qy     1589 GCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTT 1648
               :::      :::                         |||     |||
Db      975 erSerThrSerSerSerLeuThrThrAlaLeuThrGluIleThrPro-PheSerTyrIle 994

Qy     1649 TCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTG---------CTCTGCCGAGCGATG 1699
            |||||||||         |||        :::        :::     |||
Db      995 SerLeuProSerThrThrProCysProGlyThrIleThrIleThrIleValProAlaSer 1014

Qy     1700 CCAACTTCACCATGCGTA-----------------TGAGCTACGGCTCCA--------- 1732
            ||||||  |||||||||                   ||||||:::|||
Db     1015 ProThrAspProCysValGluMetAspProSerThrGluAlaThrSerProProThrThr 1034

Qy     1732 ------------------------------------------------------------ 1732

Db     1035 ProLeuThrValPheProPheThrThrGluMetValThrCysProThrSerIleSerIle 1054

Qy     1732 ------------------------------------------------------------ 1732

Db     1055 GlnThrThrLeuThrThrTyrMetAspThrSerSerMetMetProGluSerGluSerSer 1074

Qy     1733 --------------TCAAGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATA 1777
                          |||                 |||||||||     |||:::
Db     1075 IleSerProAsnAlaSerSerSerThrGlyThrGlyThrValProThrAsnThrValPhe 1094

Qy     1778 CGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGG 1837
                     |||                                   ||||||
Db     1095 ThrSerThrArg----------------------------------LeuPro------ 1100

Qy     1838 AGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACG---GTC 1894
                  |||  |||       :::     ::::::::  :::|||        |||
Db     1101 ---ThrSerGluThrTrpLeuSerAsnSerSerValIleProLeuProLeuProGlyVal 1119

Qy     1895 AGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCG 1954
            |||:::|||  |||:::   :::  :::::::|||                     |||
Db     1120 SerThrIleProLeuThrMetLysProSerSerSer------------------LeuPro 1133

Qy     1955 TATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGA 2014
               :::||||||:::     |||        |||      |||  :::
Db     1134 ThrIleLeuArgThrSerSerLysSerThrHisProSerProProThrThrArg------ 1151

Qy     2015 GTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACG 2074
                  :::|||     |||:::   :::   |||          |||      |||
Db     1152 --------ThrSerGluThrProValAlaThrThrGln---------ThrProThrThr 1165

Qy     2075 TTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
               :::|||      |||     :::::::::   :::::::::::|||
Db     1166 LeuThrSerArgArgThrThrArgIleThrSerGlnMetThrThrGlnSer 1182
```

RESULT 9
AAW20056

```
ID   AAW20056 standard; Protein; 1528 AA.
XX
AC   AAW20056;
XX
DT   10-SEP-1997  (first entry)
XX
DE   C. elegans UNC-53 protein variant 8A.
XX
KW   UNC-53; neuronal regeneration; revascularisation; wound healing;
KW   neurodegenerative disease; Alzheimer's disease; Huntingdon's;
KW   peripheral neuropathies; metastasis inhibition; cancer.
XX
OS   Caenorhabditis elegans.
XX
PN   WO9638555-A2.
XX
PD   05-DEC-1996.
XX
PF   31-MAY-1996;   96WO-EP02311.
XX
PR   31-MAY-1995;   95GB-0010944.
XX
PA   (BOGA/) BOGAERT T.
PA   (STRI/) STRINGHAM E.
PA   (VAND/) VANDEKERCKHOVE J.
XX
PI   Bogaert T,  Stringham E,  Vandekerckhove J;
XX
DR   WPI; 1997-034369/03.
DR   N-PSDB; AAT71314.
XX
PT   Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to
PT   promote neuronal regeneration, revascularisation or wound healing.
XX
PS   Claim 20; Page 106-111; 278pp; English.
XX
CC   UNC-53 protein variants 8A and 7A of Caenorhabditis elegans are new.
CC   The UNC-53 proteins and nucleic acids are useful as medicaments to
CC   promote neuronal regeneration, revascularisation or wound healing, or
CC   for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or
CC   Huntingdon's disease) or acute traumatic injuries. Transgenic cells and
CC   organisms transfected with UNC-53 cDNA can be used to determine whether
CC   a substance is an inhibitor or enhancer of the regulation of cell shape
CC   or motility or the direction of cell migration by screening for a
CC   phenotypic change in the cell. Inhibitors can be used to alleviate the
CC   spread of disease inducing cells or metastasis. Probes derived from the
CC   cDNA sequences can be used to identify homologues of the C. elegans
CC   unc-53 gene. The UNC-53 protein can be used to identify proteins which
CC   are active in the signal transduction pathway that can be used as
CC   mentioned above.
XX
SQ   Sequence   1528 AA;

Alignment Scores:
Pred. No.:               0.000139     Length:        1528
Score:                   148.50       Matches:       106
Percent Similarity:      37.48%       Conservative:  84
Best Local Similarity:   20.91%       Mismatches:    211
Query Match:             3.89%        Indels:        107
DB:                      18           Gaps:          19

US-10-008-355-1 (1-2139) x AAW20056 (1-1528)

Qy      226 ATCACAGTGTCCGATCAGGGCCTGATC--------------TTTACCAACCACCACTGC 270
            :::|||  |||:::      ||||||                  :::   |||  |||
Db        1 MetThrThrSerAsnValGluLeuIleProIleTyrThrAspTrpAlaAsnArgHisLeu 20

Qy      271 GGATACGGTGCTATC---CAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGT 327
                  |||::::::   :::|||            :::   :::|||:::
Db       21 SerLysGlySerLeuSerLysSerIleArgAspIleSerAsnAspPheArgAspTyrArg 40

Qy      328 TTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTG 387
               ||||||:::   :::          :::||||||         |||      :::
Db       41 LeuValSerGlnLeuIleAsnValIleValProIleAsnGluPheSerProAlaPheThr 60

Qy      388 CGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAG 447
            :::::::::  |||:::|||       ::::::|||  |||:::            |||
Db       61 LysArgLeuAlaLysIleThrSerAsnLeuAspGly---LeuGluThrCysLeuAspTyr 79

Qy      448 ATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCA 507
            ::::::   |||                   |||   :::|||   |||   :::
Db       80 LeuLysAsnLeuGlyLeuAsp---------CysSerLysLeuThrLysThrAspIleAsp 96

Qy      508 GACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATC 567
                  |||         :::::::::      :::      :::   |||
Db       97 SerGlyAsnLeuGlyAlaValLeuGlnLeuLeuPheLeuLeuSerThrTyrLysGlnLys 116

Qy      568 GTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAG 627
            :::      :::  ||||||  :::   :::            |||:::|||::: :::
Db      117 LeuArgGlnLeuLysLysAspGlnLysLysLeuGluGlnLeuProThrSerIle-MetPr 136
```

```
Qy    628 TTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTC 687
               :::
Db    136 oProAlaValSerLys---------------------------------------- 141

Qy    688 CGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTAT 747
                 :::|||  |||                |||||      |||||
Db    142 -------LeuProSerProArgValAlaThrSerAlaThrAlaSerAlaThrAsnProAs 159

Qy    748 AAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATG 807
            |||  |||       :::              :::||||||  |||
Db    159 nSerAsnPheProGlnMet-----------SerThrSerArgLeuGlnThr-------- 172

Qy    808 ACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGT 867
                  ||||||     :::|||  :::|||  :::|||:::
Db    173 ----ProGlnSerArgIleSerLysIleAspSerSerLysIleGlyIleLysProLysTh 191

Qy    868 A----------------------TCGAAAACGAGAACAATCCTCGTATCGAAGTTCGC 903
                                  |||  |||  :::        :::  |||
Db    191 rSerGlyLeuLysProProSerSerSerThrThrSerSerAsnAsnThrAsnSerPheAr 211

Qy    904 GGTATCAAGCAAGGCATCTGGA------------------------------------ 925
             ||||||::::::||||||
Db    211 gProSerSerArgSerSerGlyAsnAsnAsnValGlySerThrIleSerThrSerAlaLy 231

Qy    926 -------------------------------------------AGGAAGCCATGAGCGCA 942
                                                  :::|||  :::||
Db    231 sSerLeuGluSerSerSerThrTyrSerSerIleSerAsnLeuAsnArgProThrSerGl 251

Qy    943 GATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAG 1002
          |::::::  |||  |||            ::::::||||||  |||       ::
Db    251 nLeuGlnLysPro---SerArgProGlnThrGlnLeuValArgValAlaThrThrThrLy 270

Qy   1003 AATTCGATCG-----------GTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGT 1050
          :|||  |||                 ||||||||                  ||
Db    270 sIleGlySerSerLysLeuAlaAlaProLysAlaValSerThrProLysLeuAlaSerVa 290

Qy   1051 CGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCT 1110
          |   ::::::     |||:::                                 :::||
Db    290 lLysThrIleGlyAlaLysGlnGluProAspAsnSerGlyGlyGlyGlyGlyGlyMetLe 310

Qy   1111 GTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAAC 1170
          |   :::   ::::::            :::                   |||:::|||||
Db    310 uLysLeuLysLeuPheSerSerLysAsnProSerSerSerSerAsnSerProGlnProTh 330

Qy   1171 CGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCA 1230
          |                         |||:::::::::  ||||||:::       |||
Db    330 r--------------------ArgLysAlaAlaAlaValProGlnGlnGlnThrLeuSe 343

Qy   1231 CAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTT 1290
             :::                                 ||||||  |||
Db    343 rLysIleAlaAla--------------------------ProValLysSerGlyLeuLy 354

Qy   1291 GACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTC 1350
                 ::::::|||  |||  |||   :::  |||  |||
Db    354 sProProThrSerLysLeuGlySerAlaThrSerMetSerLysLeuCysThrPro---Ly 373

Qy   1351 GATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATC 1410
            :::  |||                  |||  :::||||||  |||:::       ||
Db    373 sValSerTyrArg-----------LysThrAspAlaProIleIleSerGlnGlnAspSe 389

Qy   1411 GACAAGAAATTCAAAGGCGACACGAAGAAG-------TATGCAGACTTCGTATTCGACAA 1463
            |  |||  |||||||::::::                ||||||  |||
Db    389 rLysArgCysSerLysSerSerGluGluGluSerGlyTyrAlaGlyPhe---------As 406

Qy   1464 GAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAG---TCCATGGACAAGGA 1520
            |||      |||  |||                |||
Db    406 nSerThrSerProThrSerSerSerThrGluGlySerLeuSerMetHisSerThrSerSe 426

Qy   1521 AAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCA---------GTAGAGCTTTCCAAGAG 1571
            |||  :::  :::  ||||||  |||:::          :::  |||:::  ||
Db    426 rLysSerSerThrSerAspGluLysSerProSerSerAspAspLeuThrLeuAsnAlaSe 446

Qy   1572 CGTAATAGCTGCTGCTCGC 1590
          |::::::  |||  |||
Db    446 rIleValThrAlaIleArg 452
```

```
RESULT 10
AAW20057
ID   AAW20057 standard; Protein; 1583 AA.
XX
AC   AAW20057;
XX
DT   10-SEP-1997  (first entry)
XX
DE   C. elegans UNC-53 protein variant 7A.
XX
KW   UNC-53; neuronal regeneration; revascularisation; wound healing;
KW   neurodegenerative disease; Alzheimer's disease; Huntingdon's;
KW   peripheral neuropathies; metastasis inhibition; cancer.
XX
OS   Caenorhabditis elegans.
XX
PN   WO9638555-A2.
XX
PD   05-DEC-1996.
XX
PF   31-MAY-1996;   96WO-EP02311.
XX
PR   31-MAY-1995;   95GB-0010944.
XX
PA   (BOGA/) BOGAERT T.
PA   (STRI/) STRINGHAM E.
PA   (VAND/) VANDEKERCKHOVE J.
XX
PI   Bogaert T,  Stringham E,  Vandekerckhove J;
XX
DR   WPI; 1997-034369/03.
DR   N-PSDB; AAT71315.
XX
PT   Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to
PT   promote neuronal regeneration, revascularisation or wound healing.
XX
PS   Claim 22; Page 111-116; 278pp; English.
XX
CC   UNC-53 protein variants 8A and 7A of Caenorhabditis elegans are new.
CC   The UNC-53 proteins and nucleic acids are useful as medicaments to
CC   promote neuronal regeneration, revascularisation or wound healing, or
CC   for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or
CC   Huntingdon's disease) or acute traumatic injuries. Transgenic cells and
CC   organisms transfected with UNC-53 cDNA can be used to determine whether
CC   a substance is an inhibitor or enhancer of the regulation of cell shape
CC   or motility or the direction of cell migration by screening for a
CC   phenotypic change in the cell. Inhibitors can be used to alleviate the
CC   spread of disease inducing cells or metastasis. Probes derived from the
CC   cDNA sequences can be used to identify homologues of the C. elegans
CC   unc-53 gene. The UNC-53 protein can be used to identify proteins which
CC   are active in the signal transduction pathway that can be used as
CC   mentioned above.
XX
SQ   Sequence   1583 AA;

Alignment Scores:
Pred. No.:               0.000142      Length:         1583
Score:                   148.50        Matches:        106
Percent Similarity:      37.48%        Conservative:   84
Best Local Similarity:   20.91%        Mismatches:     211
Query Match:             3.89%         Indels:         107
DB:                      18            Gaps:           19

US-10-008-355-1 (1-2139) x AAW20057 (1-1583)

Qy    226 ATCACAGTGTCCGATCAGGGCCTGATC--------------TTTACCAACCACCACTGC 270
             :::|||  |||:::       ||||||                 :::  |||  |||
Db      1 MetThrThrSerAsnValGluLeuIleProIleTyrThrAspTrpAlaAsnArgHisLeu 20
```

```
Qy     271 GGATACGGTGCTATC---CAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGT 327
                 |||::::::    :::|||           :::    :::|||:::
Db      21 SerLysGlySerLeuSerLysSerIleArgAspIleSerAsnAspPheArgAspTyrArg 40

Qy     328 TTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTG 387
              ||||||:::    :::           :::||||||        |||        :::
Db      41 LeuValSerGlnLeuIleAsnValIleValProIleAsnGluPheSerProAlaPheThr 60

Qy     388 CGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAG 447
           :::::::::     |||:::|||        ::::::|||   |||:::          |||
Db      61 LysArgLeuAlaLysIleThrSerAsnLeuAspGly---LeuGluThrCysLeuAspTyr 79

Qy     448 ATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCA 507
           ::::::   |||                      |||   :::|||   |||    :::
Db      80 LeuLysAsnLeuGlyLeuAsp---------CysSerLysLeuThrLysThrAspIleAsp 96

Qy     508 GACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATC 567
                 |||          :::::::::        :::         :::    |||
Db      97 SerGlyAsnLeuGlyAlaValLeuGlnLeuLeuPheLeuLeuSerThrTyrLysGlnLys 116

Qy     568 GTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAG 627
           :::       :::   ||||||    :::     :::           |||:::|||::: :::
Db     117 LeuArgGlnLeuLysLysAspGlnLysLysLeuGluGlnLeuProThrSerIle-MetPr 136

Qy     628 TTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTC 687
                       :::
Db     136 oProAlaValSerLys----------------------------------------- 141

Qy     688 CGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTAT 747
                   :::|||    |||                  ||||||        ||||||
Db     142 -------LeuProSerProArgValAlaThrSerAlaThrAlaSerAlaThrAsnProAs 159

Qy     748 AAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATG 807
              |||   |||          :::                 :::||||||    |||
Db     159 nSerAsnPheProGlnMet-----------SerThrSerArgLeuGlnThr-------- 172

Qy     808 ACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGT 867
                       ||||||       :::|||.   :::|||     :::|||:::
Db     173 ----ProGlnSerArgIleSerLysIleAspSerSerLysIleGlyIleLysProLysTh 191

Qy     868 A-----------------------TCGAAAACGAGAACAATCCTCGTATCGAAGTTCGC 903
                                   |||   |||    :::         :::   |||
Db     191 rSerGlyLeuLysProProSerSerSerThrThrSerSerAsnAsnThrAsnSerPheAr 211

Qy     904 GGTATCAAGCAAGGCATCTGGA------------------------------------- 925
              ||||||::::::||||||
Db     211 gProSerSerArgSerSerGlyAsnAsnAsnValGlySerThrIleSerThrSerAlaLy 231

Qy     926 ------------------------------------------AGGAAGCCATGAGCGCA 942
                                                     :::|||    :::||
Db     231 sSerLeuGluSerSerSerThrTyrSerSerIleSerAsnLeuAsnArgProThrSerGl 251

Qy     943 GATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAG 1002
           |::::::   |||   |||               ::::::||||||    |||        ::
Db     251 nLeuGlnLysPro---SerArgProGlnThrGlnLeuValArgValAlaThrThrThrLy 270

Qy    1003 AATTCGATCG-----------GTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGT 1050
           :|||   |||                    |||||||||                ||
Db     270 sIleGlySerSerLysLeuAlaAlaProLysAlaValSerThrProLysLeuAlaSerVa 290

Qy    1051 CGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCT 1110
           |    ::::::         |||:::                                :::||
Db     290 lLysThrIleGlyAlaLysGlnGluProAspAsnSerGlyGlyGlyGlyGlyGlyMetLe 310

Qy    1111 GTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAAC 1170
           |    :::    ::::::            :::                  |||:::|||||
Db     310 uLysLeuLysLeuPheSerSerLysAsnProSerSerSerSerAsnSerProGlnProTh 330

Qy    1171 CGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCA 1230
           |                      |||:::::::::   ||||||:::         |||
Db     330 r--------------------ArgLysAlaAlaAlaValProGlnGlnGlnThrLeuSe 343

Qy    1231 CAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTT 1290
              :::                                              ||||||    |||
Db     343 rLysIleAlaAla-------------------------ProValLysSerGlyLeuLy 354

Qy    1291 GACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTC 1350
                 :::::::|||       |||       |||        :::    |||    |||
Db     354 sProProThrSerLysLeuGlySerAlaThrSerMetSerLysLeuCysThrPro---Ly 373

Qy    1351 GATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATC 1410
              :::   |||                       |||    :::||||||  |||:::     ||
Db     373 sValSerTyrArg-----------LysThrAspAlaProIleIleSerGlnGlnAspSe 389

Qy    1411 GACAAGAAATTCAAAGGCGACACGAAGAAG-------TATGCAGACTTCGTATTCGACAA 1463
              |   |||    ||||||::::::                  ||||||    |||
Db     389 rLysArgCysSerLysSerSerGluGluGluSerGlyTyrAlaGlyPhe---------As 406

Qy    1464 GAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAG---TCCATGGACAAGGA 1520
              |||      |||    |||                          |||
Db     406 nSerThrSerProThrSerSerSerThrGluGlySerLeuSerMetHisSerThrSerSe 426

Qy    1521 AAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCA---------GTAGAGCTTTCCAAGAG 1571
              |||    :::     :::    ||||||    |||:::                :::   |||:::   ||
Db     426 rLysSerSerThrSerAspGluLysSerProSerSerAspAspLeuThrLeuAsnAlaSe 446

Qy    1572 CGTAATAGCTGCTGCTCGC 1590
              |::::::    |||   |||
Db     446 rIleValThrAlaIleArg 452
```

RESULT 11
ABG06375

```
ID   ABG06375 standard; Protein; 2570 AA.
XX
AC   ABG06375;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #6366.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS70562.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 36734; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
```

```
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2570 AA;

Alignment Scores:
Pred. No.:              0.000223       Length:         2570
Score:                  148.00         Matches:        133
Percent Similarity:     35.67%         Conservative:   81
Best Local Similarity:  22.17%         Mismatches:     195
Query Match:            3.87%          Indels:         191
DB:                     22             Gaps:           29

US-10-008-355-1 (1-2139) x ABG06375 (1-2570)

Qy     491 CCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACA 550
           |||   |||   ::::::|||  ||||||     |||
Db    1864 ProThrLysAsnValLysThrThrThrAsnCysPheSer--------------------- 1876

Qy     551 ACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTC 610
           :::||||||         :::::::::      :::|||
Db    1877 SerAsnThrArgLysMetThrSerLeuLeuGluLysThr--------------------- 1889

Qy     611 CCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGG 670

Db    1889 ------------------------------------------------------------ 1889

Qy     671 GCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACA--- 727
           :::   :::  |||:::       :::      |||   :::      |||  |||
Db    1890 SerLeuThrAsnTyrAlaThrSerLeuAsnThrProValSerTyrProProTrpThrPro 1909

Qy     728 ---GCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCT 784
           :::   |||:::       ::::::  |||          |||:::        :::|||
Db    1910 SerSerAlaThrLeuProSerLeuThrSerPheValTyrSerProHisSerThrGluAla 1929

Qy     785 ACAAGGCTGACG------ACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCT 838
                    |||      |||  |||   |||      |||:::  |||
Db    1930 GluIleSerThrProLysThrSerPro---ProProThrSerGlnMetValGluPhe--- 1947

Qy     839 ACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAG 898
                 :::||||||          |||   |||:::   :::|||::::::
Db    1948 ---ProValLeuGlyThrArgMetThrSerSerAsnThrGlnProLeuLeuMetThrSer 1966

Qy     899 TTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCG---CAGATCAGGCTACCC 955
           :::   :::   :::            :::   |||   :::      |||   :::|||
Db    1967 TrpAsnIleProThrAlaGluGlySerGlnPheProIleSerThrThrIleAsnValPro 1986

Qy     956 GTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTA 1015
              ||||||          :::            |||   |||   :::   |||
Db    1987 ThrSerAsnGluMetGluThrGluThrLeuHisLeuValProGlyProLeu---SerThr 2005

Qy    1016 TGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCAT 1075
              ||||||
Db    2006 PheThrAla--------------------------------------------------- 2008

Qy    1076 TCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTC 1135
           ||||||||||||                                         |||:::
Db    2009 SerGlnThrGly-----------------------------------------LeuVal 2014

Qy    1136 TCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGA 1195
           ||||||   ::::::                |||
Db    2015 SerLysAspValMetAlaMetSerSerIlePro--------------------------- 2025

Qy    1196 CGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAA 1255
              |||   :::                           ||||||   |||   |||
Db    2026 MetSerGlyIle-------------------------LeuProAsnHisGlyLeuSer 2036

Qy    1256 ATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGT---------ACAAAG 1306
                    :::      |||       :::|||::::::                |||
Db    2037 GluAsnProSerLeuSerThrSerLeuArgAlaIleThrSerThrLeuAlaAspValLys 2056

Qy    1307 ACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGC 1366
              |||      :::   |||               |||                  |||
Db    2057 HisThrPheGluLysMetThrThrSerValThrPro--------------------Gly 2069

Qy    1367 GTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAG 1426
              :::||||||:::                                         |||
Db    2070 ThrThrLeuProSerIleLeu---------------------------------SerGly 2078

Qy    1427 GCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACA 1486
              ||||||    |||:::    :::                   |||    |||
Db    2079 AlaThrSerGlySerValIleSerLysSerProIleLeuThrTrpLeuLeu--------- 2095

Qy    1487 AGTTCCATGCCATGCTCAAGT---CCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGA 1543
           ||||||:::|||        |||   |||   |||        :::    |||      :::
Db    2096 SerSerLeuProSerGlySerProProAlaThrValSerAsnAlaProHis--ValMetT 2115

Qy    1544 AAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCG 1603
                     ||||||:::|||||||||                :::     |||   |||
Db    2115 hrSerSerThrValGluValSerLysSerThrPheLeuThrSerAspMetIleSerAlaH 2135

Qy    1604 ATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTC------------------ 1645
                     |||             |||      |||::::::
Db    2135 isProPheThrAsnLeu-ThrThrLeuProSerAlaThrMetSerThrIleLeuThrArg 2154

Qy    1646 ------------------------------------------TTTTCTTTGCCGGTT 1660
                                                       |||||||||:::
Db    2155 ThrIleProThrProThrLeuGlyGlyIleThrThrGlyPheProThrSerLeuProMet 2174

Qy    1661 TGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGA 1720
              :::       |||   |||::::::          :::   |||   |||
Db    2175 SerIleAsnValThrAspAspIleVal------TyrIleSerThrHisPro--------- 2189

Qy    1721 GCTACGGCTCCATCAAGG---------GATATGAACCGCAGGACGGTG------------ 1759
                 |||   ||||||               |||   |||||||||
Db    2190 ---GluAlaSerSerArgThrThrIleThrAlaAsnProArgThrValSerHisProSer 2208

Qy    1760 ----------------------CCTGGTACAACTATCATACGACAGGCAAGGGCGTAT 1795
                                    |||   ||||||
Db    2209 SerPheSerArgLysThrMetSerProSerThrThrAsp-------------------- 2221

Qy    1796 TGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCG------ 1849
                 :::      :::   ||||||   ||||||         |||:::|||
Db    2222 ---HisThrLeuSerValGlyAlaMetProLeuProSerSerThrIleThrSerSerTrp 2240

Qy    1850 ------------ACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGC 1897
                       |||:::|||:::|||   |||::::::            |||:::
Db    2241 AsnArgIleProThrAlaSerSerProSerThrLeuIleIleProLysProThrLeuAsp 2260

Qy    1898 TCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTAT 1957
           |||:::   |||:::         ||||||||||||          |||   |||   :::
Db    2261 SerLeu---LeuAsnIleMetThrThrThrSer------ThrValProGlyAlaSerPhe 2277

Qy    1958 TCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTG 2017
              :::                               ::::::||||||
Db    2278 ProLeu------------------------------IleSerThrGlyValThrTyrPro 2287

Qy    2018 GTGACATCGAGT---TCGAACCCGATCTGCAGC---GCACAATCAGCGTGGACATCCGCT 2071
              |||::::::   |||:::||||||   |||      ::::::   |||      :::
Db    2288 PheThrAlaThrValSerSerProIleSerSerPhePheGluThrThrTrpLeuAspSer 2307

Qy    2072 ACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
           |||   |||         |||         ::::::|||   ::::::||||||
Db    2308 ThrProSerPheLeuSerThrGluAlaSerThrSerProThrAlaThrLysSer 2325
```

```
RESULT 12
ABP27518
ID   ABP27518 standard; Protein; 1368 AA.
XX
AC   ABP27518;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Streptococcus polypeptide SEQ ID NO 4212.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS   Streptococcus pyogenes.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB04789.
XX
PR   27-OCT-2000; 2000GB-0026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J,  Masignani V,  Margarit Ros YI,  Grandi G,  Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   N-PSDB; ABN68149.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein -
XX
PS   Claim 1; Page 3578; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence   1368 AA;

Alignment Scores:
Pred. No.:              0.0002        Length:          1368
Score:                  146.50        Matches:         167
Percent Similarity:     31.74%        Conservative:    105
Best Local Similarity:  19.49%        Mismatches:      228
Query Match:            3.84%         Indels:          357
DB:                     23            Gaps:            43

US-10-008-355-1 (1-2139) x ABP27518 (1-1368)

Qy    103 AATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTC 162
              |||:::||||||   |||   :::   ::::::      ||||||   |||
Db    609 AsnGluGluAsnGluAspIleLeuGluAspIleValLeuThrLeuThrLeuPheGlu--- 627

Qy    163 TACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACC 222
                   |||:::      |||         :::      :::
Db    628 ---------AspArgGluMetIleGluGluArgLeuLysThrTyrAlaHis--------- 641

Qy    223 GGTATCACAGTGTCCGATCAGGGCCTG------ATCTTTACCAACCACCACTGCGGATAC 276
                         |||:::   :::        :::               :::  |||:::
Db    642 ------LeuPheAspAspLysValMetLysGlnLeuLysArgArgArgTyrThrGlyTrp 659

Qy    277 GGTGCTATC------------------------------CAGAGCCAAAGCACGGTG 303
          |||   :::                              ||||||    |||:::
Db    660 GlyArgLeuSerArgLysLeuIleAsnGlyIleArgAspLysGlnSerGlyLysThrIle 679

Qy    304 GATCACGACTATCTGCGC---GATGGTTTCGTTTCTCGC-------------------- 339
             |||:::|||:::   |||||||||   :::|||
Db    680 ---LeuAspPheLeuLysSerAspGlyPheAlaAsnArgAsnPheMetGlnLeuIleHis 698

Qy    340 -----------ACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCC----------- 375
                     |||    |||::::::        :::|||
Db    699 AspAspSerLeuThrPheLysGluAspIleGlnLysAlaGlnValSerGlyGlnGlyAsp 718

Qy    376 ------------------------------------------GTGAAGTATCTGCGC 390
                                                    |||   :::
Db    719 SerLeuHisGluHisIleAlaAsnLeuAlaGlySerProAlaIleLysLysGlyIleLeu 738

Qy    391 AAGATCGTGAAGGTAACGGAC---------AAGGTAGAAGGACAG-------------- 426
          :::  |||||||||   |||         ||||||  |||:::
Db    739 GlnThrValLysValValAspGluLeuValLysValMetGlyArgHisLysProGluAsn 758

Qy    427 ---------------------------------CTCAAGGGTATCACTGACGAGATG 450
                                          ||||||       :::
Db    759 IleValIleGluMetAlaArgGluAsnGlnThrThrGlnLysGlyGlnLysAsnSerArg 778

Qy    451 GAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAA---------AAAGAA 501
          ||||||:::::::::   :::|||     :::||||||              ||||||
Db    779 GluArgMetLysArgIleGluGluGlyIleLysGluLeuGlySerGlnIleLeuLysGlu 798

Qy    502 AATGCAGACGAGAAC---CAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATAC 558
          :::      ||||||   ||||||                        ||||||
Db    799 HisProValGluAsnThrGlnLeu-----------------------GlnAsnGluLys 810

Qy    559 TTCCTCATCGTCTAC--------------GATGTATTCAAGGACGTTCGTATGGTATTT 603
             :::   |||              |||::::::   |||       :::
Db    811 LeuTyrLeuTyrTyrLeuGlnAsnGlyArgAspMetTyrValAspGlnGluLeuAspIle 830

Qy    604 GCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGT 663

Db    830 ------------------------------------------------------------ 830

Qy    664 CACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAA 723
                                                    ||||||   ::::::
Db    831 ------------------------------------------AsnArgLeuSerAsp 835

Qy    724 TACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGC 783
          |||    |||:::                                 :::    |||
Db    836 TyrAspValAspHis------------------------------IleValProGlnSer 845

Qy    784 TACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTC 843
          :::    ||||||                          |||   |||      :::
Db    846 PheLeuLysAspAsp--------------------------SerIleAspAsnLysVal 856

Qy    844 ACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGC 903
             :::         ::::::|||   :::::::::|||    |||       ||||||
Db    857 LeuThrArgSerAspLysAsnArgGlyLysSerAspAsnValProSerGluGluValVal 876

Qy    904 GGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAA 963
                :::      |||::::::   ::::::|||         |||:::   |||
Db    877 LysLysMetLysAsnTyrTrpArgGlnLeuLeuAsnAlaLysLeuIleThrGlnArgLys 896

Qy    964 TATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGC 1023
          :::                      |||      |||             |||
Db    897 Phe--------------------AspAsnLeuThrLysAlaGluArgGly--------- 906
```

```
Qy   1024 GGTCTCGCTCGTCTTGACGTGATAGGT---------CGTAAGCGTGCCGAGGAAAGAGCA 1074
          ||||||:::  ||||||     |||         |||:::    |||  |||
Db    907 GlyLeuSerGluLeuAspLysAlaGlyPheIleLysArgGlnLeuValGluThrArgGln 926

Qy   1075 TTCGCAGACTGGATC-----------------CGTAAGAACGGCAAG------------ 1104
                  :::                     |||  |||  |||
Db    927 IleThrLysHisValAlaGlnIleLeuAspSerArgMetAsnThrLysTyrAspGluAsn 946

Qy   1104 ------------------------------------------------------------ 1104

Db    947 AspLysLeuIleArgGluValLysValIleThrLeuLysSerLysLeuValSerAspPhe 966

Qy   1104 ------------------------------------------------------------ 1104

Db    967 ArgLysAspPheGlnPheTyrLysValArgGluIleAsnAsnTyrHisHisAlaHisAsp 986

Qy   1105 ---------AGTGCTGTCTATGGCGATGTATTG-----------TCTTCTCTCGAAAAG 1143
                   :::||||||  |||    |||                ||||||
Db    987 AlaTyrLeuAsnAlaValValGlyThrAlaLeuIleLysLysTyrProLysLeuGluSer 1006

Qy   1144 GCTTATAAGGAAGGA-------------------------------GCCAAGGCCAAC 1170
             :::     |||                                ||||||:::
Db   1007 GluPheValTyrGlyAspTyrLysValTyrAspValArgLysMetIleAlaLysSerGlu 1026

Qy   1171 CGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCA 1230
          :::|||:::     :::        |||    :::   :::::   |||
Db   1027 GlnGluIleGlyLysAlaThrAlaLysTyrPhePheTyrSerAsnIleMetAsnPhePhe 1046

Qy   1231 CAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCG--- 1287
          :::          ||||||                        ::::::::::::
Db   1047 LysThrGluIleThrLeuAlaAsnGlyGluIleArgLysArgProLeuIleGluThrAsn 1066

Qy   1288 -----------------CTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGT 1329
                           :::  ||||||  :::|||:::      :::      |||
Db   1067 GlyGluThrGlyGluIleValTrpAspLysGlyArgAspPhe------AlaThrValArg 1084

Qy   1330 AAGGTGCTGCCC---GCCATGCTCGATATTGTACGCCGGCGT----------------- 1368
          |||||||||          ::::::||||||::::::
Db   1085 LysValLeuSerMetProGlnValAsnIleValLysLysThrGluValGlnThrGlyGly 1104

Qy   1369 -----------------ATCCCT---------GCCGACAAGCTCCCCGATATATTCAAG 1401
                           :::|||         :::|||||||||
Db   1105 PheSerLysGluSerIleLeuProLysArgAsnSerAspLysLeu-------------- 1119

Qy   1402 AATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGAC 1461
                |||   :::|||         |||  |||||||||          ||||||
Db   1120 ------IleAlaArgLysLysAspTrpAspProLysLysTyrGlyGly------PheAsp 1135

Qy   1462 AAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCC----------------------- 1497
                |||  ||||||            |||
Db   1136 SerProThrValAlaTyrSerValLeuValValAlaLysValGluLysGlyLysSerLys 1155

Qy   1498 ATGCTCAAGTCC--------------------------ATGGACAAGGAAAAGTTTGCC 1530
             |||||||||                            |||::::::   |||
Db   1156 LysLeuLysSerValLysGluLeuLeuGlyIleThrIleMetGluArgSerSerPheGlu 1175

Qy   1531 ---------------------------------AAGGCTATCGAGAAAGATCCGGCA 1554
                                           |||   ::::::||||||
Db   1176 LysAsnProIleAspPheLeuGluAlaLysGlyTyrLysGluValLysLysAspLeuIle 1195

Qy   1555 GTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCC 1614
          ::::::|||  |||
Db   1196 IleLysLeuProLys-----------------------------------------Tyr 1201

Qy   1615 AATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTAC 1674
          :::   :::   :::|||  |||::::::          |||          |||:::
Db   1202 SerLeuPheGluLeuGluAsnGlyArgLysArgMetLeuAlaSerAlaGlyGluLeuGln 1221

Qy   1675 CCCGGACGT------GCTCTGCCGAGCGAT---GCCAACTTCACCATGCGTATGAGC--- 1722
             |||          ||||||||||||         ||||||         |||
Db   1222 LysGlyAsnGluLeuAlaLeuProSerLysTyrValAsnPheLeuTyrLeuAlaSerHis 1241

Qy   1723 TACGGCTCCATCAAGGGATATGAACCGCAGGACGGT-----------GCCTGGTACAAC 1770
             |||      :::||||||      |||:::|||                  :::
Db   1242 TyrGluLysLeuLysGly---SerProGluAspAsnGluGlnLysGlnLeuPheValGlu 1260

Qy   1771 TATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCC 1830
             |||         |||      |||::::::           :::      ||||||:::
Db   1261 GlnHis---------LysHisTyrLeuAspGluIleIleGluGlnIleSerGluPheSer 1277

Qy   1831 GTACAG------------------------------------------------------ 1836
             :::
Db   1278 LysArgValIleLeuAlaAspAlaAsnLeuAspLysValLeuSerAlaTyrAsnLysHis 1297

Qy   1837 --------------------------GAGAATATCCTCGACCTCTTCCGCACCAAAAAC 1869
                                    |||||||||:::    ||||||          |||
Db   1298 ArgAspLysProIleArgGluGlnAlaGluAsnIleIleHisLeuPheThrLeuThrAsn 1317

Qy   1870 TATGGT-------------------------------------------CGCTATGCC 1884
             |||                                               ||||||
Db   1318 LeuGlyAlaProAlaAlaPheLysTyrPheAspThrThrIleAspArgLysArgTyrThr 1337

Qy   1885 GAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGC 1935
                   ::::::      |||   |||   :::     |||||||||
Db   1338 SerThrLysGluValLeuAspAlaThrLeuIleHisGlnSerIleThrGly 1354
```

RESULT 13
ABB69806

```
ID   ABB69806 standard; Protein; 1795 AA.
XX
AC   ABB69806;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 36210.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL13909.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 36210; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
SQ   Sequence   1795 AA;

Alignment Scores:
Pred. No.:                0.000739       Length:          1795
Score:                    141.50         Matches:         137
Percent Similarity:       34.70%         Conservative:    109
Best Local Similarity:    19.32%         Mismatches:      303
Query Match:              3.70%          Indels:          160
DB:                       22             Gaps:            25

US-10-008-355-1 (1-2139) x ABB69806 (1-1795)

Qy     91 CTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCTCCCG 150
Db    414 LeuAsnAspIleAsnLysTyrGlnTyrLysArgTyr----------------------- 425

Qy    151 TTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTTCGGT 210
Db    426 ---------ThrTyrGlyThrAspLysAsnAspValThrGluAlaProGluIle---Lys 441

Qy    211 GGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCACTGC 270
Db    442 SerProLeuLysGlyLeuHisLeuSerGluAsnIleValIle-LeuProGluThrThrTh 461

Qy    271 GGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGTTTC 330
Db    461 rThrThrThrThrThr---------ThrLysProValValLeuThrCysProThrIleSe 478

Qy    331 GTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGC 390
Db    478 r----------------------------------------ProProAsp-------- 481

Qy    391 AAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATG 450
Db    482 ------------------ThrThrProLysProSerThrThrThrAlaValThr----- 493

Qy    451 GAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAA---------AAGAA 501
Db    494 ------------------------------LysSerThrProLysIleSerSerThrGl 503

Qy    502 AATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTC 561
Db    503 uGlnHisSerThrThrThrAlaLysThrThrThrThrLysArgProThrThrValThrGl 523

Qy    562 CTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTA 621
Db    523 uLysThrSerSer--------------------------------------------- 527

Qy    622 GGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACA------CGGGCGAC 675
Db    528 ----AlaThrGluLysProArgThrThrValValThrThrThrThrGlnLysArgSerTh 546

Qy    676 TTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGG--------------CC 720
Db    546 rThrThrHisAsnThrSerProAspThrLysThrThrIleArgSerThrThrLeuSerPr 566

Qy    721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCG---------TA 771
Db    566 oLysThrThrThrThrProSerThrThrThrProSerThrThrThrProSerThrThrTh 586

Qy    772 TCCATGCAAGGCTACAAGGCTGACGACTATGCCA---------TGACCATCGGTTTCCCG 822
Db    586 rProSerThrThrThrProSerThrThrThrProSerThrThrThrProSerThrThrTh 606

Qy    823 GGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAAC 882
Db    606 rThrValLysValSerThrHisArgProArgThrThrSerGlnLysThrThrThrAlaSe 626

Qy    883 AATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCA 942
Db    626 rThrThrThrLysLysThrThrThrSerProLysThrThr------------------- 639

Qy    943 GATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAG 1002
Db    640 ----------LysThrThrAspIleProThrSerThrThrSerLysLeuSerThrThrTh 656

Qy   1003 AATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCC 1062
Db    656 rGlnLysThrThrThrThrThrHisLysPheThrAlaAlaThrThrSerThrGluLysPr 676

Qy   1063 GAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGAT 1122
Db    676 oLysThrThrThrGluLysThrSerThrValSerThrThr------------------- 689

Qy   1123 GTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACT 1182
Db    690 ------------ThrLysLysSerThrGluSerSerProLysProThrSerSer----- 703

Qy   1183 TATTTGAGCGAGACGCTCTTCGGTGGTACCG-----------AGGTGGTTCGTTTTGCA 1230
Db    704 -------ThrGlyLysProThrThrThrProLysProSerThrArgThrThrProThrTh 721

Qy   1231 CAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTT 1290
Db    721 rThrLysValThrThrThrThrGlnIleThrThrThrThrProLeuArgSerSerThrGl 741

Qy   1291 GACGACAAGTACAAAG---------ACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCC 1341
Db    741 uThrThrSerThrGlnProProThrThrThrThrProGlnProThrThrThrThrThrLe 761

Qy   1342 GCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAG 1401
Db    761 uThrValThrProLysThrSerThrThrThrThrThrThrGluLysProIleThrSer-- 780

Qy   1402 AATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGAC 1461
Db    781 -------SerProLysProThrThrThrThrGlnLysThrThrSerThrAlaProAsnTh 798

Qy   1462 AAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAA 1521
Db    798 rThr---------LysValAlaIleThrThrGlnLysGluThrThrProThrGlnSerTh 815

Qy   1522 AAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCT 1581
Db    815 rSerThrThrIlePheThrArgLysThrThrThrAsnAsn------------------- 828

Qy   1582 GCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAG 1641
Db    829 ------------------------------ProGluProThrSerThrGluLysProIl 838

Qy   1642 CGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCC 1701
Db    838 eThrSerThrThrProLysProSerThrThrThrProLysThrSerThrValAlaSerSe 858

Qy   1702 AACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCC 1761
Db    858 rThrGluLysThrThr---IleSerSerProLysProThrThrGluLysSerThrGluAs 877

Qy   1762 TGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGAT 1821
Db    877 nProThrThr-------------------------AsnSerValLysThrSerAlaLe 888

Qy   1822 GAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCA----------------- 1864
Db    888 uThrSerSerThrGlnArgAlaThrSerThrThrSerGluProThrLysThrThrGlnAs 908

Qy   1865 ----AAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920
Db    908 nIleThrThrThrThrProLysProThrThrLeuLysThr---SerThrGlnGluAlaTh 927

Qy   1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
Db    927 rThrSerThrGlnLysValSerThrValThrIleThrThrLysLysAlaThrGluSerSe 947

Qy   1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
```

```
Db      947 rProLeuThrThrLeuSerThrGluGlu---------ProAsnThrThrProLysProLe 964

Qy     2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100
              :   :::                 ||||||    |||   ::::::   :::|||
Db      964 uArgThrThrThrProThrThrThrSerValThrAlaThrThrArgIleThrThrThrTh 984

Qy     2101 CAGTGCCCCCGTCTCATCCAAGAGC 2125
                 :::      |||:::::::::
Db      984 rIleSerGluSerSerThrGluThr 992

RESULT 14
AAY31745
ID   AAY31745 standard; Protein; 430 AA.
XX
AC   AAY31745;
XX
DT   22-NOV-1999  (first entry)
XX
DE   Mycobacterium tuberculosis specific DNA-encoded polypeptide.
XX
KW   Tuberculosis; infection; diagnosis; DNA probe.
XX
OS   Mycobacterium tuberculosis.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 4
FT                   /note= "encoded by TGA"
FT   Misc-difference 6
FT                   /note= "encoded by TGA"
FT   Misc-difference 20
FT                   /note= "encoded by TGA"
FT   Misc-difference 29
FT                   /note= "encoded by TGA"
FT   Misc-difference 54
FT                   /note= "encoded by TGA"
FT   Misc-difference 64
FT                   /note= "encoded by TGA"
FT   Misc-difference 69
FT                   /note= "encoded by TGA"
FT   Misc-difference 89
FT                   /note= "encoded by TGA"
FT   Misc-difference 99
FT                   /note= "encoded by TGA"
FT   Misc-difference 114
FT                   /note= "encoded by TGA"
FT   Misc-difference 119
FT                   /note= "encoded by TGA"
FT   Misc-difference 129
FT                   /note= "encoded by TGA"
FT   Misc-difference 159
FT                   /note= "encoded by TGA"
FT   Misc-difference 169
FT                   /note= "encoded by TGA"
FT   Misc-difference 182
FT                   /note= "encoded by TAG"
FT   Misc-difference 185
FT                   /note= "encoded by TGA"
FT   Misc-difference 219
FT                   /note= "encoded by TGA"
FT   Misc-difference 259
FT                   /note= "encoded by TGA"
FT   Misc-difference 269
FT                   /note= "encoded by TGA"
FT   Misc-difference 291
FT                   /note= "encoded by TGA"
FT   Misc-difference 323
FT                   /note= "encoded by TGA"
FT   Misc-difference 339
FT                   /note= "encoded by TGA"
FT   Misc-difference 349
FT                   /note= "encoded by TGA"
FT   Misc-difference 356
FT                   /note= "encoded by TGA"
FT   Misc-difference 366
FT                   /note= "encoded by TGA"
FT   Misc-difference 410
FT                   /note= "encoded by TGA"
FT   Misc-difference 421
FT                   /note= "encoded by TGA"
XX
PN   EP945462-A1.
XX
PD   29-SEP-1999.
XX
PF   25-MAR-1998;   98EP-0302287.
XX
PR   25-MAR-1998;   98EP-0302287.
XX
PA   (COUL ) CSIR COUNCIL SCI IND RES.
XX
PI   Kumar D,  Srivastava BS,  Srivastava R;
XX
DR   WPI; 1999-530042/45.
DR   N-PSDB; AAX87940.
XX
PT   New nucleic acid molecules, useful for detecting and identifying
PT   Mycobacterium tuberculosis
XX
PS   Disclosure; Page 15-27; 42pp; English.
XX
CC   The present sequence represents an amino acid sequence deduced
CC   from the Mycobacterium tuberculosis specific DNA fragment provided
CC   in AAX87940.  This DNA fragment comprises a StuI-StuI fragment of
CC   M. tuberculosis genomic DNA and contains an insertion sequence-like
CC   element and repetitive sequences.  The DNA fragment is useful as a
CC   probe, especially for detecting or identifying M. tuberculosis in
CC   clinical isolates and body fluids e.g. sputum, cerebrospinal fluid,
CC   pleural fluid, urine, gastric lavage, bronchial lavage, pericardial
CC   or lymph node aspirate (all claimed).  It is also useful for
CC   restriction fragment length polymorphism analysis of M. tuberculosis
CC   isolates (claimed).  The probe provides rapid and specific diagnosis
CC   of tuberculosis and M. tuberculosis infection.
XX
SQ   Sequence   430 AA;

Alignment Scores:
Pred. No.:              0.000595       Length:         430
Score:                  138.00         Matches:        133
Percent Similarity:     23.61%         Conservative:   28
Best Local Similarity:  19.50%         Mismatches:     169
Query Match:            3.49%          Indels:         352
DB:                     20             Gaps:           37

US-10-008-355-1 (1-2139) x AAY31745 (1-430)

Qy     1960 CGAATACGGGGCTACCGGAGTTACCGC------CCGTGATGTCGTTGTTCGATAGGAAAG 1907
              |||   |||         |||       |||       |||    ||||||      |||
Db       10 ArgArgArgSerLeuArgCysValArgGln***ProSerAlaArgCysLeuTrpGly*** 29

Qy     1906 CGATATGGAGCTGACCGTTCTCGGCATAGCGACCATAGTTTTGGTGCGGAAGAGGTCGA 1847
                            ||||||          :::                      |||
Db       30 IleGlySerGlyAspArgGluIleAspGlyGlu-----------------SerGlyPro 43

Qy     1846 GGATATTCTCCTGTACGGCAAACTCATCGCTCTTAGGATCCTGCTTCTCCAATACGCCCT 1787
             ||||||    |||         ::::::|||                        ||||||
Db       44 GlyTyrArgProProProArgSerHis-----------------Cys***SerArgPro 57

Qy     1786 TGCCTGTCG------TATGATAGTTGTACCAGGCACCGTCCTGCGGTTCATATCCCTTGA 1733
             |||               |||      |||         |||
Db       58 CysArgLysArgArgTyr***ArgCysArgAsp***ArgSerArgCys----------- 73

Qy     1732 TGGAGCCGTAGCTCATACGCATGGTGAAGTTGGCATCGCTCGGCAGAGCACGTCCGGGGT 1673
             |||           |||:::         |||                 |||    |||
Db       74 TrpCysHisProGlyTyrSerCysArgTyrTrpCys-----------HisProArg*** 89

Qy     1672 ACA--------------------------------------------------------- 1670
```

```
            :::
Db       90 SerCysArgCysTrpCysTyrArgArg***SerArgArgCysCysSerCysProCysCys 109

Qy     1669 ---TCTCACGCAAACCGGCAAAGAAAAGACGCTTGCCCTTCTCAATGGCATAGGCAT--- 1616
               |||          :::|||:::      ||||||       .   :::|||
Db      110 ArgSerProCys***SerArgArgArg***GlyCysProCysCysSerCysGlnHis*** 129

Qy     1615 ---------------------------------TGGCCATCGCATCGGCCTGAATAGCGC 1589
                                             :::||||||   |||
Db      130 GlyCysArgTyrCysArgTyrProGlySerArgTyrProSerSerArg------------ 145

Qy     1588 GAGCAGCAGCTATTACGCTCTTGGAAAGCTCTACTGCCGGATCTTTCTCGATAGCCTTGG 1529
                                                |||   |||
Db      146 ---------------------------------CysProSerLeuArg---------Cys 151

Qy     1528 CAAACTTTTCCTTGTCCA--------------------TGGACTTGAGCATGGCATGGA 1490
            :::    |||   ||||||                     |||                |||
Db      152 ArgArgPheArgCysProArg***ArgCysGlnArgTyrTrpCysProAsn***ThrGly 171

Qy     1489 ACTTGT---------------------------CGCTATAAGGAACCACACTCTTGTCGA 1457
               |||                            :::|||      |||      ||||||
Db      172 ArgCysCysArgCysProSerSerSerArg***GlnTyr***SerProAlaGlyCysArg 191

Qy     1456 ATACGAAGTCTGCATACTTCTTCGTGTCGCCTTTGAATTTCTTGTCGATTACATTCTTGA 1397
                        |||:::   ||||||         |||
Db      192 -----------ArgThrAlaArgCysArg---------CysCys---------------- 199

Qy     1396 ATATATCGGGGAGCTTGTCGGCAGGGATACGCCGGCGTACAATATCGAGCATGGCGGGCA 1337
                        |||                          |||   ||||||
Db      200 -----------CysCys--------------------------ArgCysTrpArgThr 206

Qy     1336 GCACCTTACGGTCGAGCGAGGGGAGGTAGTCTTTGTACTTGTCGTCAAGCGATTTGAGGA 1277
                                          |||   ||||||
Db      207 --------------------------ArgCysCysCysCysArg--------------- 212

Qy     1276 TACCGGCATGAGCATCAGGATTTGTAGCCAATGCGTTGGCAAACTGTGCAAAACGAACCA 1217
                                             ||||||::::::
Db      213 ---------------------------------CysTrpGlnSerLeuGly***SerArg 221

Qy     1216 CCTCGGTACCACCGAAGAGCGTCTCGCTCAAATAAGTCATCTCACGGTTGGCCTTGGCTC 1157
            ||||||                  |||||||||    :::   |||
Db      222 ProArg---------------SerArgSerArgArgCysSerArg--------------- 231

Qy     1156 CTTCCTTATAAGCCTTTTCGAGAGAAGACAATACATCGCCATAGACAGCACTCTTGCCGT 1097
                                             |||     :::         ||||||
Db      232 ---------------------------------ArgArgPheGlnAsnArgCysCysArg 240

Qy     1096 TCTTACGGATCCAGTCTGCGAATGCTC---TTTCCTCGGCACGCTTACGACCTATCACGT 1040
            |||   |||         :::|||  .    ||||||
Db      241 SerArgGlyPheArgIleArgCysCysSerPhePro------------------------ 252

Qy     1039 CAAGACGAGCGAGACCGCGGTTCATACCGATCGAATTCTTCCAATAGTTAGCACTCTGAG 980
                                     :::|||:::                |||
Db      253 --------------------GlyPheArgAsnArg---------------His***Ile 260

Qy      979 CATACTTGCTGGCATATTTGATACGGGTAGCCTGATCTGCGCTCATGGCTTCCTTCCAGA 920
               |||:::|||
Db      261 LeuArgCysPheHis--------------------------------------------- 265

Qy      919 TGCCTTGCTTGATACCGCGAACTTCGATACGAGGATTGTTCTCGTTTTCGATACGATCTT 860
                                     ||||||         |||   |||   |||
Db      266 --------------------CysArgTyr***SerCysArgArgCysArg--------- 275

Qy      859 CCACACCCCAAGAAGTGAGGTAGCGATCCGTACTGCCCGGGAAACCGATGGTCATGGCAT 800

Db      275 ------------------------------------------------------------ 275

Qy      799 AGTCGTCAGCCTTGTAGCCTTGCATGGATACGGCAGCGAAGTAAACGGGCTTATAGGGTT 740

Db      275 ------------------------------------------------------------ 275

Qy      739 TATTGTCCTTGCTGTATTCGGCCGGCCGGTTGTCGGCACCGGCATACACGCGGAATACGC 680
               ||||||   |||               |||||||||
Db      276 ---CysProArgCysPheGlyCysArgGlyCysArg----------------------- 286

Qy      679 TGAAGTCGCCCGTGTGACGCGGCCACATCCAGTTGTCCGTATCGCCTCCGAACTTACCTA 620
                        |||         |||          :::|||    |||
Db      287 -----------CysGlnGlyCys***SerHisArgArgPheArgCysArgGluCysCys 302

Qy      619 CAGAGCTGGGAGGAGCAAATACCATACGAACGTCCTTGAATACATCGTAGACGATGAGGA 560

Db      302 ------------------------------------------------------------ 302

Qy      559 AGTATTCGTTGTTGGAATAGAAAGGCTCTACGATGCAGAGTTGGTTCTCGTCTGCATTTT 500
            :::   |||||||||               |||||||||     ||||||
Db      303 AsnCysArgCysTrp---------------ArgCysArgGluCysSerArgArgProGly 317

Qy      499 CTT--------------------------------------------------------- 497
            |||
Db      318 LeuProGlyArgAsp***ArgProValGlyHisArgLysIleProThrCysCysPheArg 337

Qy      496 -----------------TTTTGGCCAGTTCTTGGCATACCTCCTGAGCTTTGCGCAGAC 455
                                |||   |||      ||||||      |||
Db      338 Cys***ArgSerProArgSerArgProAlaLeu***TrpProProGlySerCys***Thr 357

Qy      454 GCTCCATCTCGTCAGTGATACCCTTGAGCTGTCCTTCTACCTTGTCCGTTACCTTCACGA 395
                                                              |||   |||
Db      358 Asn---------------------------------------------------ProIleArg 361

Qy      394 TCTTGCGCAGATACTTCACGGAAAGACCCGGAATCGGAAGCTCCTCACCCATCGTGCGAG 335
               |||         ||||||   |||                |||
Db      362 CysCysProSer***SerArgProIlePro-----------AlaArgProArgLeuPro 377

Qy      334 AAACGAAACCATCGCGCAGATAGTCGTGATCCACCGTGCTTTGGCTCTGGATAGCACCGT 275
             |||::::::|||          ||||||
Db      378 GlyArgSerTyrArgTrp------------ProPro----------------------- 385

Qy      274 ATCCGCAGTGGTGGT-----------TGGTAAAGATCAGGCCCTGATCGGACACTGTGA 227
               :::||||||            |||   ||||||       ||||||
Db      386 ThrLysSerGlyArgSerGlnAsnCysTrpHisArgSerSerGlySerArgThr------ 403

Qy      226 TACCGGTACATCCGCCACCGA------AGATAACCACGGCATTGGCAATGGACGGCTTGT 173
               ||||||  |||||||||      |||   |||         |||::::::
Db      404 ---ArgTyr---ArgHisArgCys***ArgIleProThrThrAlaHisTrpSerSer*** 421

Qy      172 CGAAAC 167
             |||:::
Db      422 ArgSer 423
```

RESULT 15
ABB61705
ID   ABB61705 standard; Protein; 684 AA.
XX
AC   ABB61705;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 11907.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;

```
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL05808.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 11907; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   684 AA;

Alignment Scores:
Pred. No.:              0.000823        Length:         684
Score:                  138.00          Matches:        151
Percent Similarity:     34.85%          Conservative:   109
Best Local Similarity:  20.24%          Mismatches:     264
Query Match:            3.49%           Indels:         222
DB:                     22              Gaps:           36

US-10-008-355-1 (1-2139) x ABB61705 (1-684)

Qy   2067 GATGTCCACGCTGATTGTGCGCTGCAGATCGGGTTCGAACTCGATGTCACCACTCATAGC 2008
          |||:::|||  |||           :::     |||:::  |||      |||
Db      3 AspLeuHisValAspGluLeuGluValValAspIleGluValValValMetLeuHisArg 22

Qy   2007 TTCCCAGTTGCCATCGAAAGCAAGACCGATCAGACGGCCGTTCTTATCGAATACGGGGCT 1948
                     :::|||   :::|||:::       |||   :::
Db     23 ---------ProValGluGlyGluThrGluGlyGlnGlyValHisValAlaGlyThrVal 39

Qy   1947 ACCGGAGTTACCGCCCGTGATGTCGTTGTTCGATAGGAAAGCGATATGGAGCTGACCGTT 1888
          :::        |||        :::|||                 ||||||
Db     40 SerHisAlaAlaAlaAspHisArgIleValAla------------SerGluLeuGlyAsn 55

Qy   1887 CTCGGCATAGCGACCATAGTTTTTGGTGCGGAAGAGGTCGAGGATATT------CTCCTG 1834
          ||||||:::        |||        ||||||:::|||  :::     ||||||
Db     56 LeuGlyLeuValValArgValAspHisValGluGluLeuGluHisLeuAlaAlaLeuLeu 75

Qy   1833 TACGGCAAA-----------------------CTCATCGCTCTTAGGATCCTGCTTCTC 1798
                :::                        :::   ||||||:::::   |||
Db     76 AlaAlaArgGlnHisHisSerValGlyHisAlaAlaMetGlyLeuArgLeuIleGlnLeu 95

Qy   1797 CAATACGCCCTTGCCTGTCGTATGATAGTTGTACCAGGCACCGTC---------CTGCGG 1747
          :::                 :::::|||     |||   |||        |||
Db     96 Glu-----------------LeuValValThrValGlyGlyValHisGlnLeuLeuVal 109

Qy   1746 TTCATATCCCTTGATGGAGCCGTAGCTCATACGCATGGTGAAGTTGGCATCGCTCGGCAG 1687
             :::   :::  ||||||:::  |||     |||||||||
Db    110 GluLeuGlnValGlyGlyAlaIleValHisProGluGlyGluVal--------------- 124

Qy   1686 AGCACGTCCGGG-----------------------------GTACATCTCACGCAAACC 1657
             |||:::|||                           :::||||||
Db    125 ---ThrAlaGlyAlaLeuAspValValAsnAlaArgLeuGlnMetHisLeuGluValLeu 143

Qy   1656 GGCAAAGAAAAGACGCTTGCCCTTCTC---------AATGGCATAGGCATTGGCCATCGC 1606
                |||:::  |||   ::::::       :::   :::   :::
Db    144 GlnLeuGluGluValLeuValValValLeuGlnTrpAspLeuValSerValLeuIleVal 163

Qy   1605 ATCGGCCTGAATAGCGCGAGCAGCAGCTATTACGCTCTTGGAAAGCTCTACTGCCGGATC 1546
          :::|||   |||          :::
Db    164 LeuGlySerAsnGluAspLeuHisAsn--------------------------------- 172

Qy   1545 TTTCTCGATAGCCTTGGCAAACTTTTCCTTGTCCATGGACTTGAGCATGGCATGGAACTT 1486
                        |||   |||  ||||||  |||         |||  ||||||
Db    173 --------------GlyIleLeuSerLeuValArgGlyHisValGlnGly---GluLeu 186

Qy   1485 GTCGCTATAAGGAACCACACTCTTGTCGAATACGAAGTCTGCATACTTCTTCGTGTCGCC 1426
                :::           :::||||||    :::      |||     :::|||
Db    187 GluGlyLeu-----------ValValGluThrLeuLeuAlaGlyLeuAspHisIleAla 202

Qy   1425 TTTGAATTTCTTGTCGATTACATTCTTGAATATATC-----------GGGGAGCTTGTC 1378
                  :::|||           |||   :::           |||  |||
Db    203 GlySerLeuValValGlyValProAlaGluGlyValGlnArgIleGlyGlySerLeuAsn 222

Qy   1377 GGC--------------------AGGGATACGCCGGCGTACAATATCGAGCATGGCGGG 1339
          |||                      |||              :::|||  ||||||
Db    223 GlyGlyLeuHisIleAspGlyIleAlaAspValArgValGluValLeuGluValGlyGly 242

Qy   1338 CAG-----------------CACCTTACGGTCGAGCGAGGGGAGGTAGTCTTTGTACTT 1297
          |||                   :::   |||:::     |||   |||  :::|||
Db    243 GlnThrPheGlyHisValAspGlyIleAspValAspThrLeuGluGlyValIleLeuLeu 262

Qy   1296 GTCGTCAAGCGATTT-----------------------GAGGATACCGGCATGAGCATC 1261
          ||||||  |||                            |||   |||:::
Db    263 ValValLeuArgIleGlyLeuHisGlnMetProValValAlaAspAsnGlyLeuValTyr 282

Qy   1260 AGGATTTGTAGCCAATGCGTTGGCAAACTGTGCAAAACGAACCACCTCGGTACCACCGAA 1201
             :::     :::                         |||   |||
Db    283 LeuValGluAspArgIleAla--------------------HisArgGlyValValVal 295

Qy   1200 GAGCGTCTCGCTCAAATAAGTCATCTCACGGTTGGCCTTGGCTCCTTCCTTATAAGCCTT 1141
                  |||:::   :::  :::|||:::  :::         ::::::::|||
Db    296 HisGluLysAlaArgGluAlaLeuValThrMetAlaIle---------MetValAlaLeu 312

Qy   1140 TTCGAGAGAAGACAATACATCGCCATAGACAGCACTCTTGCCGTTCTTACGGATCCAGTC 1081
                                :::   :::  :::   :::|||        |||
Db    313 ValHisAsnGluAlaValGlnLeuValLysThrAlaValGluLeuLeuValHisAlaVal 332

Qy   1080 TGCGAATGCTCTTTCCTCGGCACGCTTACGACCTATCACGTCAAGACGAGCGAGACCGCG 1021
                                              :::        :::   |||
Db    333 ----------------------------------ArgIleAlaLeuAspAspValAla 340

Qy   1020 GTTCATACCGATCGAATTCTTCCAATAGTTAGCACTCTGAGCATACTTGCTGGCATATTT 961
             |||  |||   :::        :::     |||   ::::::     :::
Db    341 GlnHisIleAspGlyLeu------HisIleGluLeuLeuValLeuValAspProValHis 358

Qy    960 GATACGGGTAGCCTGATCTGCGCTCATGGCTTCCTTCCAGATGCCTTGCTTGATACCGCG 901
          ||||||        :::     |||      :::  |||   |||:::
Db    359 AspThr---------LeuArgValHisLeuHisMetLeuAspValLeuVal--------- 372

Qy    900 AACTTCGATACGAGGATTGTTCTCGTTTTCGATACGATCTTCCACACCCCAAGAAGTGAG 841
                        :::||||||| :::   :::           |||
Db    373 --------------ValValLeuValValGluAlaValValIleLeuPro--------- 384

Qy    840 GTAGCGATCCGTACTGCCCGGGAAACCGATGGTCATGGCATAGTCGTCAGCCTTGTAGCC 781
          |||   :::                   |||:::  :::|||
Db    385 ValProLeu--------------------GlyTyrAlaValVal--------------- 392

Qy    780 TTGCATGGATACGGCAGCGAAGTAAACGGGCTTATAGGGTTTATTGTCCTTGCTGTATTC 721
                     |||::::::|||  |||:::  ||||||
Db    393 -----------GlyAlaGlnValThrGlyIleAlaGlyPhe------------------ 402

Qy    720 GGCCGGCCGGTTGTCGGCACCGGCATACACGCGGAATACGCTGAAGTCGCCCGTGTG--- 664
                        |||     ||||||   |||        :::|||  :::
Db    403 --------------GlyTrpHisIleHisValGluSerGlyHisLeuAlaGluIleVal 417

Qy    663 --------------------ACGCGGCCACATCCAGTTGTCCGTATCGCCTCCGAACTT 625
                              |||             |||   |||     ||||||
Db    418 GlyGlnThrLeuGlnValProThrGlyGlnGlnLeuProValAspIleGlnValGluLeu 437

Qy    624 ACCTACAGAGCTGGGAGGAGCAAATACCATACGAACGTCCTTGAATACATCGTAGACGAT 565
          |||                                          :::|||  |||
Db    438 Thr-----------------------------------AlaThrIleLeuValLysAsp 445
```

```
Qy    564 GAGGAAGTATTCGTTGTTGGAATAGAAAGGCTCTACGATGCAGAGTTGGTTCTCGTCTGC 505
                :::      :::      ||||||:::         |||:::|||:::
Db    446 MetLeuLeuAlaAlaLeuArgProGluArgIleValGlnIleGluValValValGlu--- 464

Qy    504 ATTTTCTTTTTTGGCCAGTTCTTGGCATACCTCCTGAGCTTTGCGCAGACGCTCCATCTC 445
                      |||               |||            |||   :::   :::
Db    465 ------------GlyIleGluAlaAsnIleLeuGlnValValGlyGlnIleValIleVal 480

Qy    444 GTCAGTGATACCCTTGAGCTGTCCTTCTACCTTGTCCGTTACCTTCACGATCTTGCGCAG 385
          |||      |||:::   :::         :::||||||      |||   :::
Db    481 ValGlnGlyThrValHisValArg------IleValArgThrAspHisHisValValAsp 498

Qy    384 ATACTTCACGGAAAGACCCGGAATCGGAAGCTCCTCACCCATCGTGCGAGAAACGAAACC 325
          ::::::|||
Db    499 ValValHis---------------------------------------------------Leu 502

Qy    324 ATCGCGCAGATAGTCGTGATCCAC-------------------CGTGCTTTGGCTCTG 286
          :::      :::      :::|||                     :::   ||||||:::
Db    503 LeuAspMetMetGluAlaValHisPheValValLeuSerProPheGlnCysLeuAlaIle 522

Qy    285 GATAGCACC------GTATCCGCAGTGGTGGTTGGTAAAGATCAGGCCCTGATCGGACAC 232
          :::            |||      ||||||      :::|||   :::|||
Db    523 GluValAlaProTrpValAspLeuValValGlyArgGlnAspAspSerLeu--------- 539

Qy    231 TGTGATACCGGTACATCCGCCACCGAAGATAACCACGGCATTGGCAATGGACGGCTTGTC 172
                      |||      |||   :::   ||||||:::            ||||||
Db    540 ---HisIleIleThrGluGlnThrGlyAsnGlyHisGlyValHisGluGlnLeuLeuVal 558

Qy    171 ------------------------GAAACTGTAGAGCGAATCCAACGGGAGCGTAAAGCC 136
                                  :::   :::|||   :::            :::
Db    559 TyrValLeuAlaGlyProValAspAspValLeuGluGlyLeuProValLeuGlyGluVal 578

Qy    135 GAGCTCACGCATTCGATCCAGATTCTCCTGATTGAGTTCGTT------GAGGAGCCACAT 82
          |||   ||||||            :::|||:::      |||            ||||||
Db    579 GluGluThrHisGlnThrHisHisIleLeuLeuValLeuValValHisHisArgProHis 598

Qy     81 GCCTTTGTCGGCTTTGGC 64
                :::|||   |||
Db    599 HisValIleGlyGluGly 604
```

Search completed: May 16, 2003, 12:28:32
Job time : 165 secs

DP

```
                GenCore version 5.1.4_p5_4578
          Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:         May 16, 2003, 12:23:56 ; Search time 30.5 Seconds
                                          (without alignments)
                                          4126.927 Million cell updates/sec

Title:          US-10-008-355-1
Perfect score:  3820
Sequence:       1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters:      525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110339_9386/app_query.fasta_1.2311
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_33_@runat_16052003_110339_9386 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Issued_Patents_AA:*
                1:  /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
                2:  /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
                3:  /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
                4:  /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
                5:  /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                6:  /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                               SUMMARIES
                %
Result          Query
  No.   Score   Match Length DB ID                   Description
 -----------------------------------------------------------------------------
 c  1     138    3.5    430  4  US-09-156-836B-2      Sequence 2, Appli
    2     130    3.4    883  2  US-08-953-492-2       Sequence 2, Appli
 c  3     129    3.3    430  3  US-08-997-897-2       Sequence 2, Appli
    4   127.5    3.3   1194  2  US-08-488-940-18      Sequence 18, Appl
    5   123.5    3.2    713  4  US-09-059-584-53      Sequence 53, Appl
    6   122.5    3.2    800  2  US-08-488-940-4       Sequence 4, Appli
    7   122.5    3.2    813  2  US-08-488-940-3       Sequence 3, Appli
    8   122.5    3.2   1181  2  US-08-488-940-2       Sequence 2, Appli
    9   122.5    3.2   1194  2  US-08-488-940-1       Sequence 1, Appli
   10   122.5    3.2   1194  2  US-08-488-940-17      Sequence 17, Appl
 c 11     119    3.0   2152  4  US-09-036-987A-3      Sequence 3, Appli
 c 12     119    3.0   2152  4  US-09-370-700-3       Sequence 3, Appli
   13   115.5    3.0    846  2  US-07-728-215-33      Sequence 33, Appl
   14   115.5    3.0    846  4  US-08-938-085A-33     Sequence 33, Appl
   15     115    3.0    885  1  US-08-484-105-14      Sequence 14, Appl
   16     115    3.0    885  1  US-08-484-106-14      Sequence 14, Appl
   17     112    2.9    652  1  US-08-765-081-6       Sequence 6, Appli
   18     112    2.9    652  3  US-09-098-082-6       Sequence 6, Appli
   19     112    2.9   1014  4  US-09-078-347A-3      Sequence 3, Appli
   20   111.5    2.9   1013  4  US-08-860-886-2       Sequence 2, Appli
   21     111    2.9   1114  4  US-08-811-583-2       Sequence 2, Appli
   22     110    2.9   1027  4  US-08-446-137B-2      Sequence 2, Appli
   23     109    2.9   1577  2  US-08-793-824-2       Sequence 2, Appli
   24     106    2.8    972  3  US-08-335-844A-23     Sequence 23, Appl
   25     104    2.7    888  4  US-09-134-001C-3032   Sequence 3032, Ap
   26   103.5    2.7    562  4  US-09-211-704A-4      Sequence 4, Appli
   27   103.5    2.7    933  2  US-08-682-847-2       Sequence 2, Appli
   28   102.5    2.7    725  2  US-08-816-105A-1      Sequence 1, Appli
   29   102.5    2.7    833  4  US-09-514-302-3       Sequence 3, Appli
   30   102.5    2.7    856  4  US-09-134-001C-5438   Sequence 5438, Ap
   31   102.5    2.7   1938  4  US-09-514-302-2       Sequence 2, Appli
 c 32   102.5    2.6   3898  2  US-08-876-991-2       Sequence 2, Appli
 c 33   102.5    2.6   3898  2  US-09-059-853-2       Sequence 2, Appli
   34     102    2.7    413  4  US-09-134-001C-5111   Sequence 5111, Ap
 c 35     102    2.6   1172  1  US-08-313-288B-19     Sequence 19, Appl
   36     102    2.7   1940  2  US-08-644-271-30      Sequence 30, Appl
   37     102    2.7   1940  4  US-09-077-955-34      Sequence 34, Appl
   38   101.5    2.7    747  1  US-07-854-596B-40     Sequence 40, Appl
   39   101.5    2.7   3025  6  5223423-3            Patent No. 5223423
   40     101    2.6    433  1  US-08-417-492-2       Sequence 2, Appli
   41     101    2.6    772  4  US-09-134-078-28      Sequence 28, Appl
   42   100.5    2.6    371  4  US-09-199-637A-295    Sequence 295, App
   43   100.5    2.6    638  1  US-08-712-241-6       Sequence 6, Appli
   44     100    2.6    571  6  5200183-17           Patent No. 5200183
   45     100    2.6    668  1  US-08-445-050-9       Sequence 9, Appli

                               ALIGNMENTS

RESULT 1
US-09-156-836B-2
; Sequence 2, Application US/09156836B
; Patent No. 6242585
; GENERAL INFORMATION:
;  APPLICANT:  Srivastava, Ranjana
;  APPLICANT:  Kumar, Deepak
;  APPLICANT:  Srivastava, Brahm Shanker
;  TITLE OF INVENTION:  MYCOBACTERIUM TUBERCULOSIS SPECIFIC DNA FRAGMENT
;  FILE REFERENCE:  U 011876-4
;  CURRENT APPLICATION NUMBER: US/09/156,836B
;  CURRENT FILING DATE:  1998-09-18
;  PRIOR APPLICATION NUMBER:  08/997,897
;  PRIOR FILING DATE:  1997-12-24
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 430
;   TYPE: PRT
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (4)
;   OTHER INFORMATION: amino acid has not been identified
;   NAME/KEY: UNSURE
;   LOCATION: (6)
;   OTHER INFORMATION: amino acid has not been identified
;   NAME/KEY: UNSURE
;   LOCATION: (20)
;   OTHER INFORMATION: amino acid has not been identified
;   NAME/KEY: UNSURE
;   LOCATION: (29)
;   OTHER INFORMATION: amino acid has not been identified
;   NAME/KEY: UNSURE
;   LOCATION: (54)
;   OTHER INFORMATION: amino acid has not been identified
```

```
;    NAME/KEY: UNSURE
;    LOCATION: (64)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (69)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (89)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (99)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (114)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (119)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (129)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (159)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (169)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (182)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (185)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (219)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (259)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (269)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (291)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (323)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (339)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (349)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (356)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (366)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (410)
;    OTHER INFORMATION: amino acid has not been identified
;    NAME/KEY: UNSURE
;    LOCATION: (421)
;    OTHER INFORMATION: amino acid has not been identified
US-09-156-836B-2

Alignment Scores:
Pred. No.:              7.39e-05      Length:         430
Score:                  138.00        Matches:        133
Percent Similarity:     23.61%        Conservative:   28
Best Local Similarity:  19.50%        Mismatches:     169
```

```
Query Match:            3.49%         Indels:         352
DB:                     4             Gaps:           37

US-10-008-355-1 (1-2139) x US-09-156-836B-2 (1-430)

Qy   1960 CGAATACGGGGCTACCGGAGTTACCGC------CCGTGATGTCGTTGTTCGATAGGAAAG 1907
           |||   |||     |||     |||     |||    ||||||      |||
Db     10 ArgArgArgSerLeuArgCysValArgGln***ProSerAlaArgCysLeuTrpGly*** 29

Qy   1906 CGATATGGAGCTGACCGTTCTCGGCATAGCGACCATAGTTTTTGGTGCGGAAGAGGTCGA 1847
                       ||||||         :::                    |||
Db     30 IleGlySerGlyAspArgGluIleAspGlyGlu------------------SerGlyPro 43

Qy   1846 GGATATTCTCCTGTACGGCAAACTCATCGCTCTTAGGATCCTGCTTCTCCAATACGCCCT 1787
           ||||||   |||      ::::::|||                       ||||||
Db     44 GlyTyrArgProProProArgSerHis------------------Cys***SerArgPro 57

Qy   1786 TGCCTGTCG------TATGATAGTTGTACCAGGCACCGTCCTGCGGTTCATATCCCTTGA 1733
           |||               |||     |||        |||
Db     58 CysArgLysArgArgTyr***ArgCysArgAsp***ArgSerArgCys------------ 73

Qy   1732 TGGAGCCGTAGCTCATACGCATGGTGAAGTTGGCATCGCTCGGCAGAGCACGTCCGGGGT 1673
           |||        |||:::          |||                 |||   |||
Db     74 TrpCysHisProGlyTyrSerCysArgTyrTrpCys------------HisProArg*** 89

Qy   1672 ACA--------------------------------------------------------- 1670
           :::
Db     90 SerCysArgCysTrpCysTyrArgArg***SerArgArgCysCysSerCysProCysCys 109

Qy   1669 ---TCTCACGCAAACCGGCAAAGAAAAGACGCTTGCCCTTCTCAATGGCATAGGCAT--- 1616
              |||            :::|||:::       ||||||         :::|||
Db    110 ArgSerProCys***SerArgArgArg***GlyCysProCysCysSerCysGlnHis*** 129

Qy   1615 --------------------------------TGGCCATCGCATCGGCCTGAATAGCGC 1589
                                           :::||||||   |||
Db    130 GlyCysArgTyrCysArgTyrProGlySerArgTyrProSerSerArg------------ 145

Qy   1588 GAGCAGCAGCTATTACGCTCTTGGAAAGCTCTACTGCCGGATCTTTCTCGATAGCCTTGG 1529
                                               |||   |||
Db    146 --------------------------------CysProSerLeuArg---------Cys 151

Qy   1528 CAAACTTTTCCTTGTCCA--------------------TGGACTTGAGCATGGCATGGA 1490
           :::   |||  ||||||                      |||               |||
Db    152 ArgArgPheArgCysProArg***ArgCysGlnArgTyrTrpCysProAsn***ThrGly 171

Qy   1489 ACTTGT---------------------------CGCTATAAGGAACCACACTCTTGTCGA 1457
              |||                             :::|||      |||     ||||||
Db    172 ArgCysCysArgCysProSerSerSerArg***GlnTyr***SerProAlaGlyCysArg 191

Qy   1456 ATACGAAGTCTGCATACTTCTTCGTGTCGCCTTTGAATTTCTTGTCGATTACATTCTTGA 1397
                       |||:::   ||||||         |||
Db    192 ------------ArgThrAlaArgCysArg---------CysCys--------------- 199

Qy   1396 ATATATCGGGGAGCTTGTCGGCAGGGATACGCCGGCGTACAATATCGAGCATGGCGGGCA 1337
                       |||                           |||   ||||||
Db    200 ------------CysCys--------------------------ArgCysTrpArgThr 206

Qy   1336 GCACCTTACGGTCGAGCGAGGGGAGGTAGTCTTTGTACTTGTCGTCAAGCGATTTGAGGA 1277
                                                 |||   ||||||
Db    207 --------------------------ArgCysCysCysCysArg-------------- 212

Qy   1276 TACCGGCATGAGCATCAGGATTTGTAGCCAATGCGTTGGCAAACTGTGCAAAACGAACCA 1217
                                                 ||||||::::::
Db    213 --------------------------------CysTrpGlnSerLeuGly***SerArg 221

Qy   1216 CCTCGGTACCACCGAAGAGCGTCTCGCTCAAATAAGTCATCTCACGGTTGGCCTTGGCTC 1157
           ||||||                ||||||||| :::   |||
Db    222 ProArg---------------SerArgSerArgArgCysSerArg--------------- 231

Qy   1156 CTTCCTTATAAGCCTTTTCGAGAGAAGACAATACATCGCCATAGACAGCACTCTTGCCGT 1097
                                                |||   :::        ||||||
Db    232 --------------------------------ArgArgPheGlnAsnArgCysCysArg 240
```

```
Qy   1096 TCTTACGGATCCAGTCTGCGAATGCTC---TTTCCTCGGCACGCTTACGACCTATCACGT 1040
          |||  |||     :::|||      ||||||
Db    241 SerArgGlyPheArgIleArgCysCysSerPhePro------------------------ 252

Qy   1039 CAAGACGAGCGAGACCGCGGTTCATACCGATCGAATTCTTCCAATAGTTAGCACTCTGAG 980
                                 :::|||:::              |||
Db    253 --------------------GlyPheArgAsnArg--------------His***Ile 260

Qy    979 CATACTTGCTGGCATATTTGATACGGGTAGCCTGATCTGCGCTCATGGCTTCCTTCCAGA 920
                |||:::|||
Db    261 LeuArgCysPheHis--------------------------------------------- 265

Qy    919 TGCCTTGCTTGATACCGCGAACTTCGATACGAGGATTGTTCTCGTTTTCGATACGATCTT 860
                                 ||||||     |||  |||  |||
Db    266 --------------------CysArgTyr***SerCysArgArgCysArg--------- 275

Qy    859 CCACACCCCAAGAAGTGAGGTAGCGATCCGTACTGCCCGGGAAACCGATGGTCATGGCAT 800

Db    275 ------------------------------------------------------------ 275

Qy    799 AGTCGTCAGCCTTGTAGCCTTGCATGGATACGGCAGCGAAGTAAACGGGCTTATAGGGTT 740

Db    275 ------------------------------------------------------------ 275

Qy    739 TATTGTCCTTGCTGTATTCGGCCGGCCGGTTGTCGGCACCGGCATACACGCGGAATACGC 680
             ||||||  |||         |||||||||
Db    276 ---CysProArgCysPheGlyCysArgGlyCysArg------------------------ 286

Qy    679 TGAAGTCGCCCGTGTGACGCGGCCACATCCAGTTGTCCGTATCGCCTCCGAACTTACCTA 620
                      |||       |||        :::|||  |||
Db    287 -----------CysGlnGlyCys***SerHisArgArgPheArgCysArgGluCysCys 302

Qy    619 CAGAGCTGGGAGGAGCAAATACCATACGAACGTCCTTGAATACATCGTAGACGATGAGGA 560

Db    302 ------------------------------------------------------------ 302

Qy    559 AGTATTCGTTGTTGGAATAGAAAGGCTCTACGATGCAGAGTTGGTTCTCGTCTGCATTTT 500
          :::  |||||||||               |||||||||     ||||||
Db    303 AsnCysArgCysTrp---------------ArgCysArgGluCysSerArgArgProGly 317

Qy    499 CTT--------------------------------------------------------- 497
          |||
Db    318 LeuProGlyArgAsp***ArgProValGlyHisArgLysIleProThrCysCysPheArg 337

Qy    496 ------------------TTTTGGCCAGTTCTTGGCATACCTCCTGAGCTTTGCGCAGAC 455
                            |||  |||     ||||||    |||
Db    338 Cys***ArgSerProArgSerArgProAlaLeu***TrpProProGlySerCys***Thr 357

Qy    454 GCTCCATCTCGTCAGTGATACCCTTGAGCTGTCCTTCTACCTTGTCCGTTACCTTCACGA 395
                                                           |||  |||
Db    358 Asn---------------------------------------------------ProIleArg 361

Qy    394 TCTTGCGCAGATACTTCACGGAAAGACCCGGAATCGGAAGCTCCTCACCCATCGTGCGAG 335
             |||       ||||||    |||                   |||
Db    362 CysCysProSer***SerArgProIlePro------------AlaArgProArgLeuPro 377

Qy    334 AAACGAAACCATCGCGCAGATAGTCGTGATCCACCGTGCTTTGGCTCTGGATAGCACCGT 275
             |||:::::|||         ||||||
Db    378 GlyArgSerTyrArgTrp------------ProPro----------------------- 385

Qy    274 ATCCGCAGTGGTGGT------------TGGTAAAGATCAGGCCCTGATCGGACACTGTGA 227
             :::||||||              |||   ||||||       ||||||
Db    386 ThrLysSerGlyArgSerGlnAsnCysTrpHisArgSerSerGlySerArgThr------ 403

Qy    226 TACCGGTACATCCGCCACCGA------AGATAACCACGGCATTGGCAATGGACGGCTTGT 173
             ||||||  |||||||||    |||  |||       |||::::::
Db    404 ---ArgTyr---ArgHisArgCys***ArgIleProThrThrAlaHisTrpSerSer*** 421

Qy    172 CGAAAC 167
           |||:::
Db    422 ArgSer 423
```

RESULT 2

```
US-08-953-492-2
; Sequence 2, Application US/08953492
; Patent No. 5849555
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, James
;    APPLICANT:  Jaworski, Deborah
;    APPLICANT:  Lawlor, Elizabeth
;    APPLICANT:  Wang, Min
;    TITLE OF INVENTION:  NOVEL valS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/953,492
;      FILING DATE:  17-OCT-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/844,064
;      FILING DATE:  18-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9607991.8
;      FILING DATE:  18-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P31458-4/1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  883 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-953-492-2

Alignment Scores:
Pred. No.:               0.000829     Length:          883
Score:                   130.00       Matches:         104
Percent Similarity:      32.68%       Conservative:    63
Best Local Similarity:   20.35%       Mismatches:      168
Query Match:             3.40%        Indels:          176
DB:                      2            Gaps:            24

US-10-008-355-1 (1-2139) x US-08-953-492-2 (1-883)

Qy    574 GATGTATTC-----------------------AAGGACGTTCGTATGGTATTTGCTCCT 609
          |||||||||                       |||      :::|||    |||
Db     26 AspValPheLysProSerGlyAspGlnLysAlaLysProTyrSerIleValIleProPro 45

Qy    610 CCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACG 669
          |||:::    ||||||                                        |||
Db     46 ProAsnValThrGlyLysLeu--------------------------------HisLeu 54

Qy    670 GGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGC 729
          |||                                                  ::::::
Db     55 Gly-------------------------------------------------HisAla 57

Qy    730 AAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAG 789
             |||             :::   :::                  |||||||||:::
```

```
Db       58 TrpAspThr---ThrLeuGlnAspIleIleIleArgGlnLysArgMetGlnGlyPheAsp 76

Qy      790 GCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTC---ACT 846
                                 |||:::     ||||||           :::   |||
Db       77 -----------------ThrLeuTrpLeuProGlyMetAspHisAlaGlyIleAlaThr 90

Qy      847 TCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGT 906
                     ||||||:::|||:::      |||         |||    ::::::
Db       91 GlnAlaLysValGluGluArgLeuArgGlyGluGlyIleSerArgTyrAspLeuGlyArg 110

Qy      907 -----------ATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACC 954
                           |||      ||||||:::              :::   ||||||
Db      111 GluSerPheLeuThrLysValTrpGluTrpLysAsp-----------GluTyrAlaThr 126

Qy      955 CGTATCAAATATGCC---AGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATC 1011
               ||||||         |||         |||   :::|||   :::
Db      127 ThrIleLysGluGlnTrpGlyLysMetGlyLeuSerValAspTyrSerArgGluArgPhe 146

Qy     1012 GGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGA 1071
               ::::::   ||||||:::                       :::|||      :::
Db      147 ThrLeuAspGluGlyLeuSer-----------------------LysAlaValArgLys 158

Qy     1072 GCATTCGCAGACTGGATCCGTAAG-----------AACGGCAAGAGTGCTGTCTATGGC 1119
               |||   |||      :::|||                |||:::      :::
Db      159 ValPheValAspLeuTyrLysLysGlyTrpIleTyrArgGlyGluPheIleIleAsnTrp 178

Qy     1120 GAT--------------GTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAG 1164
            |||                     ||||||   :::|||      :::|||:::
Db      179 AspProAlaAlaArgThrAlaLeuSerAspIleGluValIleHisLysAspValGluGly 198

Qy     1165 GCCAACCGTGAGATGACTTATTTG---------------------------------- 1188
            |||         |||   |||:::
Db      199 AlaPheTyrHisMetAsnTyrMetLeuGluAspGlySerArgValLeuGluValAlaThr 218

Qy     1189 ------AGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAAC 1242
                     ||||||:::||||||
Db      219 ThrArgProGluThrMetPheGlyAspVal------------------------------ 228

Qy     1243 GCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTAC 1302
            |||:::|||   ||||||:::                             |||   :::|||
Db      229 AlaValAlaValAsnProGlu--------------------------AspProArgTyr 239

Qy     1303 AAAGACTACCTCCCCTCG--------------CTCGACCGTAAGGTGCTGCCCGCCATG 1347
            ||||||   :::                  :::   |||::::::|||     :::
Db      240 LysAspLeuIleGlyLysAsnValIleLeuProIleAlaAsnLysLeuIleProIleVal 259

Qy     1348 CTCGAT-------------------------------------ATTGTACGCCGGCGT 1368
               |||                                        |||
Db      260 GlyAspGluHisAlaAspProGluPheGlyThrGlyValValLysIleThrProAlaHis 279

Qy     1369 ATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC 1428
               |||   |||   |||                  |||:::
Db      280 AspProAsnAspPheLeuValGlyGlnArgHisAsnLeuProGlnValAsnValMetAsn 299

Qy     1429 GACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAG 1488
            |||                  |||   |||:::   |||   :::         |||:::
Db      300 AspAspGlyThrMetAsnAspLeuAlaPheGluPheSerGlyMet---------AspArg 316

Qy     1489 TTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGAT 1548
            |||   |||                     |||   ||||||      |||
Db      317 PheGluAlaArg--------------LysAlaValValAlaLysLeuGluGluIleGly 331

Qy     1549 CCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCT-------------- 1593
                  |||::::::   |||   |||   :::         :::
Db      332 AlaLeuValLysIleGluLysArgValHisSerValGlyHisSerGluArgThrGlyVal 351

Qy     1594 ---------------------------------------------ATTCAGGCCGAT 1605
                                                         :::      :::
Db      352 ValValGluProArgLeuSerThrGlnTrpPheValLysMetAspGlnLeuAlaLysAsn 371

Qy     1606 GCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGT 1665
            |||:::||||||               :::|||
Db      372 AlaIleAlaAsnGlnAspThrGluAspLys--------------------------Val 382

Qy     1666 GAGATGTACCCCGGACGTGCTCTGCCGAGC--------------------------GAT 1698
            |||   ||||||   |||                  :::                  |||
Db      383 GluPheTyrProProArgPheAsnAspThrPheLeuGlnTrpMetGluAsnValHisAsp 402

Qy     1699 GCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGT 1758
               :::   ::::::   :::|||                             |||
Db      403 TrpValIleSerArgGlnLeuTrpTrpGlyHis-----------------GlnIlePro 416

Qy     1759 GCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGC 1818
            ||||||||||||                  |||:::      :::   ::::::   |||:::
Db      417 AlaTrpTyrAsn------AlaAspGlyGluMetTyrValGlyGluGluAlaProGluGly 434

Qy     1819 GATGAGTTTGCCGTACAGGAGAATATCCTCGAC 1851
            |||      :::            |||::::::||||||
Db      435 AspGlyTrpThrGlnAspGluAspValLeuAsp 445
```

```
RESULT 3
US-08-997-897-2
; Sequence 2, Application US/08997897C
; Patent No. 6114514
; GENERAL INFORMATION:
;  APPLICANT: SRIVASTAVA, RANJANA
;  APPLICANT: KUMAR, DEEPAK
;  APPLICANT: SRIVASTAVA, BRAHM SHANKER
;  TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS SPECIFIC DNA FRAGMENT
;  FILE REFERENCE: u011469-7
;  CURRENT APPLICATION NUMBER: US/08/997,897C
;  CURRENT FILING DATE: 1997-12-24
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 430
;   TYPE: PRT
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (4)
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (6)
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (20)
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (29)
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (54)
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (64)
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (69)
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (89)
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (99)
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (114)
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (119)
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (129)
;   FEATURE:
;   NAME/KEY: UNSURE
```

```
;    LOCATION: (159)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (169)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (182)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (185)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (219)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (259)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (269)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (291)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (323)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (339)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (349)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (356)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (366)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (410)
;    FEATURE:
;    NAME/KEY: UNSURE
;    LOCATION: (421)
US-08-997-897-2

Alignment Scores:
Pred. No.:               0.000589     Length:         430
Score:                   129.00       Matches:        132
Percent Similarity:      23.46%       Conservative:   28
Best Local Similarity:   19.35%       Mismatches:     170
Query Match:             3.26%        Indels:         352
DB:                      3            Gaps:           37

US-10-008-355-1 (1-2139) x US-08-997-897-2 (1-430)

Qy   1960 CGAATACGGGGCTACCGGAGTTACCGC------CCGTGATGTCGTTGTTCGATAGGAAAG 1907
          |||  |||     |||    |||      |||      ||||||      |||
Db     10 ArgArgArgSerLeuArgCysValArgGln***ProSerAlaArgCysLeuTrpGly*** 29

Qy   1906 CGATATGGAGCTGACCGTTCTCGGCATAGCGACCATAGTTTTTGGTGCGGAAGAGGTCGA 1847
                   ||||||          :::                   |||
Db     30 IleGlySerGlyAspArgGluIleAspGlyGlu-----------------SerGlyPro 43

Qy   1846 GGATATTCTCCTGTACGGCAAACTCATCGCTCTTAGGATCCTGCTTCTCCAATACGCCCT 1787
          ||||||   |||      ::::::|||                           ||||||
Db     44 GlyTyrArgProProProArgSerHis-----------------Cys***SerArgPro 57

Qy   1786 TGCCTGTCG------TATGATAGTTGTACCAGGCACCGTCCTGCGGTTCATATCCCTTGA 1733
          |||          |||     |||      |||
Db     58 CysArgLysArgArgTyr***ArgCysArgAsp***ArgSerArgCys------------ 73

Qy   1732 TGGAGCCGTAGCTCATACGCATGGTGAAGTTGGCATCGCTCGGCAGAGCACGTCCGGGGT 1673
          |||          |||:::        |||                |||   |||

Db     74 TrpCysHisProGlyTyrSerCysArgTyrTrpCys------------HisProArg*** 89

Qy   1672 ACA--------------------------------------------------------- 1670
          :::
Db     90 SerCysArgCysTrpCysTyrArgArg***SerArgArgCysCysSerCysProCysCys 109

Qy   1669 ---TCTCACGCAAACCGGCAAAGAAAAGACGCTTGCCCTTCTCAATGGCATAGGCAT--- 1616
             |||            :::|||:::     ||||||             :::|||
Db    110 ArgSerProCys***SerArgArgArg***GlyCysProCysCysSerCysGlnHis*** 129

Qy   1615 --------------------------------TGGCCATCGCATCGGCCTGAATAGCGC 1589
                                          :::||||||  |||
Db    130 GlyCysArgTyrCysArgTyrProGlySerArgTyrProSerSerArg------------ 145

Qy   1588 GAGCAGCAGCTATTACGCTCTTGGAAAGCTCTACTGCCGGATCTTTCTCGATAGCCTTGG 1529
                                           |||   |||
Db    146 --------------------------------CysProSerLeuArg---------Cys 151

Qy   1528 CAAACTTTTCCTTGTCCA--------------------TGGACTTGAGCATGGCATGGA 1490
          :::   |||   ||||||                    |||                |||
Db    152 ArgArgPheArgCysProArg***ArgCysGlnArgTyrTrpCysProAsn***ThrGly 171

Qy   1489 ACTTGT-----------------------------CGCTATAAGGAACCACACTCTTGTCGA 1457
             |||                           :::|||     |||     ||||||
Db    172 ArgCysCysArgCysProSerSerSerArg***GlnTyr***SerProAlaGlyCysArg 191

Qy   1456 ATACGAAGTCTGCATACTTCTTCGTGTCGCCTTTGAATTTCTTGTCGATTACATTCTTGA 1397
                      |||:::  ||||||          |||
Db    192 ------------ArgThrAlaArgCysArg---------CysCys--------------- 199

Qy   1396 ATATATCGGGGAGCTTGTCGGCAGGGATACGCCGGCGTACAATATCGAGCATGGCGGGCA 1337
                      |||                          |||   ||||||
Db    200 -----------CysCys--------------------------ArgCysTrpArgThr 206

Qy   1336 GCACCTTACGGTCGAGCGAGGGGAGGTAGTCTTTGTACTTGTCGTCAAGCGATTTGAGGA 1277
                                    |||   ||||||
Db    207 --------------------------ArgCysCysCysCysArg--------------- 212

Qy   1276 TACCGGCATGAGCATCAGGATTTGTAGCCAATGCGTTGGCAAACTGTGCAAAACGAACCA 1217
                                          ||||||::::::
Db    213 --------------------------------CysTrpGlnSerLeuGly***SerArg 221

Qy   1216 CCTCGGTACCACCGAAGAGCGTCTCGCTCAAATAAGTCATCTCACGGTTGGCCTTGGCTC 1157
          ||||||                  |||||||||   :::   |||
Db    222 ProArg---------------SerArgSerArgArgCysSerArg--------------- 231

Qy   1156 CTTCCTTATAAGCCTTTTCGAGAGAAGACAATACATCGCCATAGACAGCACTCTTGCCGT 1097
                                         |||   :::        ||||||
Db    232 --------------------------------ArgArgPheGlnAsnArgCysCysArg 240

Qy   1096 TCTTACGGATCCAGTCTGCGAATGCTC---TTTCCTCGGCACGCTTACGACCTATCACGT 1040
          |||   |||      :::|||         ||||||
Db    241 SerArgGlyPheArgIleArgCysCysSerPhePro----------------------- 252

Qy   1039 CAAGACGAGCGAGACCGCGGTTCATACCGATCGAATTCTTCCAATAGTTAGCACTCTGAG 980
                              :::|||:::                  |||
Db    253 --------------------GlyPheArgAsnArg---------------His***Ile 260

Qy    979 CATACTTGCTGGCATATTTGATACGGGTAGCCTGATCTGCGCTCATGGCTTCCTTCCAGA 920
              |||:::|||
Db    261 LeuArgCysPheHis--------------------------------------------- 265

Qy    919 TGCCTTGCTTGATACCGCGAACTTCGATACGAGGATTGTTCTCGTTTTCGATACGATCTT 860
                              ||||||   |||   |||   |||
Db    266 --------------------CysArgTyr***SerCysArgArgCysArg--------- 275

Qy    859 CCACACCCCAAGAAGTGAGGTAGCGATCCGTACTGCCCGGGAAACCGATGGTCATGGCAT 800

Db    275 ------------------------------------------------------------ 275

Qy    799 AGTCGTCAGCCTTGTAGCCTTGCATGGATACGGCAGCGAAGTAAACGGGCTTATAGGGTT 740

Db    275 ------------------------------------------------------------ 275
```

```
Qy    739 TATTGTCCTTGCTGTATTCGGCCGGCCGGTTGTCGGCACCGGCATACACGCGGAATACGC 680
             |||      |||            ||||||||
Db    276 ---CysArgArgCysPheGlyCysArgGlyCysArg------------------------ 286

Qy    679 TGAAGTCGCCCGTGTGACGCGGCCACATCCAGTTGTCCGTATCGCCTCCGAACTTACCTA 620
                     |||             |||         :::|||    |||
Db    287 -----------CysGlnGlyCys***SerHisArgArgPheArgCysArgGluCysCys 302

Qy    619 CAGAGCTGGGAGGAGCAAATACCATACGAACGTCCTTGAATACATCGTAGACGATGAGGA 560

Db    302 ------------------------------------------------------------ 302

Qy    559 AGTATTCGTTGTTGGAATAGAAAGGCTCTACGATGCAGAGTTGGTTCTCGTCTGCATTTT 500
          :::   ||||||||               ||||||||     ||||||
Db    303 AsnCysArgCysTrp--------------ArgCysArgGluCysSerArgArgProGly 317

Qy    499 CTT--------------------------------------------------------- 497
          |||
Db    318 LeuProGlyArgAsp***ArgProValGlyHisArgLysIleProThrCysCysPheArg 337

Qy    496 -----------------TTTTGGCCAGTTCTTGGCATACCTCCTGAGCTTTGCGCAGAC 455
                                |||   |||    ||||||     |||
Db    338 Cys***ArgSerProArgSerArgProAlaLeu***TrpProProGlySerCys***Thr 357

Qy    454 GCTCCATCTCGTCAGTGATACCCTTGAGCTGTCCTTCTACCTTGTCCGTTACCTTCACGA 395
                                                            |||   |||
Db    358 Asn-----------------------------------------------ProIleArg 361

Qy    394 TCTTGCGCAGATACTTCACGGAAAGACCCGGAATCGGAAGCTCCTCACCCATCGTGCGAG 335
             |||       ||||||  |||                    |||
Db    362 CysCysProSer***SerArgProIlePro------------AlaArgProArgLeuPro 377

Qy    334 AAACGAAACCATCGCGCAGATAGTCGTGATCCACCGTGCTTTGGCTCTGGATAGCACCGT 275
             |||::::::|||          ||||||
Db    378 GlyArgSerTyrArgTrp-----------ProPro---------------------- 385

Qy    274 ATCCGCAGTGGTGGT-----------TGGTAAAGATCAGGCCCTGATCGGACACTGTGA 227
             :::||||||              |||  ||||||        ||||||
Db    386 ThrLysSerGlyArgSerGlnAsnCysTrpHisArgSerSerGlySerArgThr------ 403

Qy    226 TACCGGTACATCCGCCACCGA------AGATAACCACGGCATTGGCAATGGACGGCTTGT 173
             ||||||  |||||||||        |||   |||         |||::::::
Db    404 ---ArgTyr---ArgHisArgCys***ArgIleProThrThrAlaHisTrpSerSer*** 421

Qy    172 CGAAAC 167
          |||:::
Db    422 ArgSer 423
```

RESULT 4
US-08-488-940-18
; Sequence 18, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,940
; FILING DATE: 09-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1194 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-488-940-18

```
Alignment Scores:
Pred. No.:              0.00188        Length:         1194
Score:                  127.50         Matches:        133
Percent Similarity:     35.05%         Conservative:   99
Best Local Similarity:  20.09%         Mismatches:     263
Query Match:            3.34%          Indels:         167
DB:                     2              Gaps:           35

US-10-008-355-1 (1-2139) x US-08-488-940-18 (1-1194)

Qy    223 GGTATCACAGTGTCCGATCAGGGCCTG------------------ATCTTTACCAACCAC 264
            |||  |||::::::   |||   |||                    |||       :::
Db    571 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 590

Qy    265 CACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGAT 324
            |||  ||||||         :::      |||::::::                |||
Db    591 HisProGlyTyrThrIleTyrGluArgAspSerSerIle---------ValThrHisAsp 607

Qy    325 GGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCG---------GGTCTTTCC 375
                :::   |||||||:::           ||||||:::
Db    608 AsnAspIlePheArgThrIle---------LeuProMetAspGlnGluPheThrTyrArg 624

Qy    376 GTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGT 435
            ||||||       :::     ::::::     |||                 |||
Db    625 ValLysAsnArgGluGlnAlaTyrArgIleAsnLysLys---------------SerGly 639

Qy    436 ATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAA 495
            :::   :::|||:::                 |||           |||   |||
Db    640 LeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLysLys 659

Qy    496 AAAGAA------------------------------------------------------ 501
             |||
Db    660 GlyGluLysProTyrAspProPheAspArgSerHisLeuLysLeuPheThrIleLysTyr 679

Qy    502 ---AATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATAC 558
             :::   |||   |||:::|||         |||         :::   :::|||
Db    680 ValAspValAspThrAsnGluLeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArg 699

Qy    559 ---------TTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCT 609
                     |||      :::||||||           ::::::::::::
Db    700 AsnLeuAspPheArgAspLeuTyrAspProArgAspLysAlaLysLeuLeuTyr------ 717

Qy    610 CCCAGCTCTGTAGGTAAGTTCGGA---------------------GGCGATACGGACAAC 648
             :::::::::      ||||||                      |||      ::::::
Db    718 ---AsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLysValGluAsp 736

Qy    649 TGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGAC 708
                       |||               :::     ||||||  |||
Db    737 -----------AsnHisAspAspThrAsnArgIleIleThrValTyrMetGly------ 750

Qy    709 AACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCC 768
             ||||||          :::|||     |||
Db    751 LysArgPro---------GluGlyGluAsnAlaSerTyrHis------------------ 761

Qy    769 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 828
                      |||    ||||||   |||                           |||
```

```
Db    762 ---------LeuAlaTyrAspAlaAspArgTyrThrGluGluGluArgGluValTyrSer 778

Qy    829 ACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCT 888
                ||||||   |||            :::   |||       ::::::||||||
Db    779 TyrLeuArgTyr---ThrGlyThrProIleProAspAsnProAsnAspLysAsnAsnSer 797

Qy    889 CGTATC------------------GAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAA 930
          ::::::                  |||   |||     |||   |||         :::
Db    798 GlnLeuValValSerValAlaGlyThrValGluGlyThrAsnGlnAspIleSerLeuLys 817

Qy    931 GCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAG------ 984
                |||      ::::|||      :::        |||      |||
Db    818 PhePheGluIleAspLeuThrSerArgProAlaHisGlyGlyLysThrGluGlnGlyLeu 837

Qy    985 ---------AGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGT 1035
                   |||      :::      :::|||      |||:::      |||   :::
Db    838 SerProLysSerLysProPheAlaThrAspSerGlyAlaMetSerHisLysLeuGluLys 857

Qy   1036 CTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAG 1095
             |||::::::      |||      :::|||:::      |||:::   :::
Db    858 AlaAspLeuLeu------LysAlaIleGlnGluGlnLeuIleAlaAsn---ValHisSer 874

Qy   1096 AAC--------------------------GGCAAGAGTGCTGTCTATGGCGATGTATTG 1128
          |||                                     |||         |||
Db    875 AsnAspAspTyrPheGluValIleAspPheAlaSerAspAlaThrIleThrAspArgAsn 894

Qy   1129 TCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTG 1188
             :::         |||   |||:::|||:::                        :::
Db    895 GlyLysValTyrPheAlaAspLysAspGlySer---ValThrLeuProThrGlnProVal 913

Qy   1189 AGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTG 1248
             |||   |||      |||         ||||||:::   :::
Db    914 GlnGluPheLeuLeuSerGlyHisValArgValArgTyrLysGlu--------------- 928

Qy   1249 GCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTAC------ 1302
                |||      :::               ||||||:::|||   :::|||
Db    929 ------LysProIleGlnAsnGlnAla------LysSerValAspValGluTyrThrVal 944

Qy   1303 -----------------------AAAGACTACCTCCCCTCGCTC---GACCGTAAGGTG 1335
                                  |||:::   |||   |||   |||   |||:::
Db    945 GlnPheThrProLeuAsnProAspAspAspPheArgProGlyLeuLysAspThrLysLeu 964

Qy   1336 CTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATA 1395
          |||         :::               |||   :::   :::|||
Db    965 LeuLysThrLeu------AlaIleGlyAspThrIleThrSerGlnGluLeuLeuAlaGln 982

Qy   1396 TTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTA 1455
          :::::::::::::::|||         |||               |||      :::
Db    983 AlaGlnSerIleLeuAsnLysAsnHisProGly-----------TyrThrIleTyrGlu 998

Qy   1456 TTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGAC 1515
             |||   |||:::|||   :::   :::                  :::      ||||||
Db    999 ArgAspSerSerIleValThrHisAspAsnAspIlePheArgThrIleLeuProMetAsp 1018

Qy   1516 AAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTA 1575
             :::|||      :::         ::::::   |||      ::::::|||
Db   1019 GlnGluPheThrTyrArgValLysAsnArgGluGlnAlaTyrArgIleAsnLysLysSer 1038

Qy   1576 ATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC---TATGCCATTGAG 1632
                                   |||   ::::::         |||   ::::::
Db   1039 GlyLeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLys 1058

Qy   1633 AAGGGCAAG---------------------------------CGTCTTTTCTTTGCCGGT 1659
             ||||||:::                                 :::||||||
Db   1059 LysGlyGluLysProTyrAspProPheAspArgSerHisLeuLysLeuPheThrIleLys 1078

Qy   1660 TTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACC--------- 1710
                ::::::                  |||   |||:::            |||
Db   1079 TyrValAspValAspThrAsnGluLeuLeuLysSerGluGlnLeuLeuThrAlaSerGlu 1098

Qy   1711 ATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCC------TGG 1764
                :::   :::      :::      |||:::|||:::|||   |||
Db   1099 ArgAsnLeuAspPheArgAspLeu------TyrAspProArgAspLysAlaLysLeuLeu 1116

Qy   1765 TACAAC--------------------------TATCATACGACAGGCAAGGGCGTATTG 1797
          ||||||                          |||   |||||||||
Db   1117 TyrAsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLys--------- 1133

Qy   1798 GAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTC 1857
                                          |||::::::|||      |||
Db   1134 --------------------------------ValGluAspAsnHisAspAspThrAsn 1142

Qy   1858 CGC------ACCAAAAACTATGGTCGC------TATGCCGAGAACGGTCAGCTCCATATC 1905
          |||      |||      |||:::            ||||||         |||:::
Db   1143 ArgIleIleThrValTyrMetGlyLysArgProGluGlyGluAsnAlaSerTyrHisLeu 1162

Qy   1906 GCTTTC 1911
          |||:::
Db   1163 AlaTyr 1164
```

```
RESULT 5
US-09-059-584-53
; Sequence 53, Application US/09059584
; Patent No. 6440701
;  GENERAL INFORMATION:
;    APPLICANT:  Myers, Lisa E
;    APPLICANT:  Schryvers, Anthony B
;    APPLICANT:  Harkness, Robin E
;    APPLICANT:  Loosmore, Sheena M.
;    APPLICANT:  Du, Run-Pan
;    APPLICANT:  Yang, Yan-Ping
;    APPLICANT:  Klein, Michel H
;    TITLE OF INVENTION:  Transferrin Receptor Genes of Moraxella
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6th Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/059,584
;      FILING DATE:  14-APR-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/778,570
;      FILING DATE:  03-JAN-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
;      REGISTRATION NUMBER:  24973
;      REFERENCE/DOCKET NUMBER:  1038-794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  713 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-059-584-53

Alignment Scores:
Pred. No.:                 0.00313     Length:         713
Score:                     123.50      Matches:        123
Percent Similarity:        31.34%      Conservative:   76
Best Local Similarity:     19.37%      Mismatches:     269
Query Match:               3.23%       Indels:         167
DB:                        4           Gaps:           24
```

```
US-10-008-355-1 (1-2139) x US-09-059-584-53 (1-713)

Qy     355 CTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAG 414
Db     156 IleProPheAspLysAsnLeuIleGluTyrLeuLysLysSerSerGluValValSerLys 175

Qy     415 GTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAG 474
Db     176 PheGluAlaGlnLysGlyGlyIleGluAsnAsnThrArgLeuThrHisLysAspLeuSer 195

Qy     475 GTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAG 534
Db     196 SerGluGlnLysGluAlaLysValLysGluAlaLeuAspAsnAlaLeu------------ 211

Qy     535 CCTTTCTATTCCAACAACGAATACTTCCTCATCGTC---------TACGATGTATTCAAG 585
Db     212 ThrGlnPheAlaGlnGluLysTyrLysGluLeuIleGluAsnAlaHisAspLysLysSer 231

Qy     586 GACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGAC 645
Db     232 AspAlaArg--------------------------------------------------- 234

Qy     646 AACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCC 705
Db     234 ------------------------------------------------------------ 234

Qy     706 GACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCT 765
Db     235 ---AsnArgAspLeuGluTyrValLysSerGlyPheAsnTyr------------------ 247

Qy     766 GCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTC------ 819
Db     248 ---------LeuSerGlyTyrThrAlaThrAspHisAspLysLysThrAsnTyrArgGly 264

Qy     820 ------------------------------------------------CCGGGCAGTACG 831
Db     265 TyrTyrGlyAlaLeuTyrTyrLysGlySerGluThrAlaLysGluLeuProGlnThrSer 284

Qy     832 GATCGCTACCTCACTTCTTGGGGT------GTGGAAGATCGTATCGAAAACGAGAACAAT 885
Db     285 AlaLysTyrLysGlyTyrTrpAspPheMetThrAspAlaThrLeuAspAsnLysTyrThr 304

Qy     886 CCTCGTATCGAAGTTCGCGGTATC---AAGCAAGGCATCTGGAAGGAAGCCATGAGCGCA 942
Db     305 ---------AspLeuProGlyIleAlaArgGlnThrGlnTrpArgSerLeuValSerThr 321

Qy     943 GATCAG---GCTACCCGTATC-----------------------AAATATGCCAGCAAG 975
Db     322 AspGluTyrAlaThrLeuLeuThrAspLysAsnAsnLysProSerAspTyrAsnGlyAla 341

Qy     976 TATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGT--------------------- 1014
Db     342 TyrGlyHisSerSerGluPheAspValAsnPheAlaAspLysLysIleLysGlyLysLeu 361

Qy    1015 ATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCA 1074
Db     362 IleSerAsnGlnLeuSerGlyThrAlaValThrAlaLysGluArgTyrLysIleGluAla 381

Qy    1075 TTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCT 1134
Db     382 AspIleHisGlyAsnArgPheArgGlySerAla---------------------ThrAla 394

Qy    1135 CTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTAT---TTGAGC 1191
Db     395 SerAspLysAla------GluAspSerLysThrGlnHisProPheThrSerAspAlaThr 412

Qy    1192 GAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCT 1251
Db     413 AsnLysLeuGluGlyGlyPheTyrGlyProLysGlyGluGluLeuAlaGlyLysPheLeu 432

Qy    1252 ACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTAC 1311
Db     433 ThrAspAspAsnLysLeuPheGlyValPheGlyAlaLysArgAspLysValGluLysThr 452

Qy    1312 CTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATC 1371
Db     453 GluAlaIleLeuAspAlaTyrAlaLeu--------------------------------- 461

Qy    1372 CCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGAC---AAGAAATTCAAAGGC 1428
Db     462 -----------------GlyThrPheAsnAsnThrAsnLysAlaThrThrPheThrPro 475

Qy    1429 GACACGAAGAAGTATGCAGAC---TTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGAC 1485
Db     476 PheThrLysLysGlnLeuAspAsnPheGlyAsnAlaLysLysLeuValLeuGlySerThr 495

Qy    1486 AAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAA 1545
Db     496 ValIleAsnLeuValSerThrAspAlaThrLysAsnGluPheThrLysLysPheThrLys 515

Qy    1546 GATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGAT 1605
Db     516 Asp-----------------LysProThrSerAlaThrAsnLysAlaGlyGluThr--- 528

Qy    1606 GCGATGGCCAATGCCTATGCCATTGAGAAG------GGCAAGCGTCTTTTCTTTGCCGGT 1659
Db     529 LeuMetValAsnAspGluValIleValLysThrTyrGlyLysAsnPheGluTyrLeuLys 548

Qy    1660 TTGCGTGAGATGTACCCCGGACGTGCT-----------CTGCCGAGCGATGCCAACTTC 1707
Db     549 PheGlyGluLeuSerValGlyAspSerHisSerValPheLeuGlnGlyGluArgThrAla 568

Qy    1708 ACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTAC 1767
Db     569 ThrThrGlyGluLysAlaValProThrThrGlyLysAlaLysTyrLeuGlyAsnTrpVal 588

Qy    1768 AACTATCATACG---------ACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGC 1818
Db     589 GlyTyrIleThrGlyAlaGlyThrGlyLysSerPheAsnGluAlaGlnAsp--------- 605

Qy    1819 GATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGC 1878
Db     606 ------------IleAlaAspPheAspIleAspPheGluArgLysSerValLysGlyLys 621

Qy    1879 TATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGC--- 1935
Db     622 LeuThrThrGlnGlyArgThrAspProValPheAsnIleLysGlyGluIleAlaGlyAsn 641

Qy    1935 ------------------------------------------------------------ 1935
Db     642 GlyTrpThrGlyLysAlaSerThrThrLysAlaAspAlaGlyGlyTyrLysIleAspSer 661

Qy    1936 GGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGAT 1995
Db     662 SerSerThrGlyLysSerIleValIleGluAsnAlaGluValThrGlyGlyPheTyrGly 681

Qy    1996 GGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
Db     682 ProAsnAlaAsnGluMetGlyGlySerPheThrHisAspThrAsp 696

RESULT 6
US-08-488-940-4
; Sequence 4, Application US/08488940
; Patent No. 5854049
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Guy L.
;    TITLE OF INVENTION:  PLASMIN-RESISTANT STREPTOKINASE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,940
;      FILING DATE:  09-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  05433/009001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  800 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-488-940-4

Alignment Scores:
Pred. No.:               0.00432       Length:          800
Score:                   122.50        Matches:         130
Percent Similarity:      33.98%        Conservative:    99
Best Local Similarity:   19.29%        Mismatches:      254
Query Match:             3.21%         Indels:          191
DB:                      2             Gaps:            35

US-10-008-355-1 (1-2139) x US-08-488-940-4 (1-800)

Qy    223 GGTATCACAGTGTCCGATCAGGGCCTG-----------------ATCTTTACCAACCAC 264
Db    177 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 196

Qy    265 CACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGAT 324
Db    197 HisProGlyTyrThrIleTyrGluArgAspSerSerIle---------ValThrHisAsp 213

Qy    325 GGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCG---------GGTCTTTCC 375
Db    214 AsnAspIlePheArgThrIle---------LeuProMetAspGlnGluPheThrTyrArg 230

Qy    376 GTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGT 435
Db    231 ValLysAsnArgGluGlnAlaTyrArgIleAsnLysLys---------------SerGly 245

Qy    436 ATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAA 495
Db    246 LeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLysLys 265

Qy    496 AAAGAA------------------------------------------------------ 501
Db    266 GlyGluLysProTyrAspProPheAspArgSerHisLeuLysLeuPheThrIleLysTyr 285

Qy    502 ---AATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATAC 558
Db    286 ValAspValAspThrAsnGluLeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArg 305

Qy    559 --------TTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCT 609
Db    306 AsnLeuAspPheArgAspLeuTyrAspProArgAspLysAlaLysLeuLeuTyr------ 323

Qy    610 CCCAGCTCTGTAGGTAAGTTCGGA--------------------GGCGATACGGACAAC 648
Db    324 ---AsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLysValGluAsp 342

Qy    649 TGGATGTGGCCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGAC 708
Db    343 -----------AsnHisAspAspThrAsnArgIleIleThrValTyrMetGly------ 356

Qy    709 AACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCC 768
Db    357 LysArgPro---------GluGlyGluAsnAlaSerTyrHisLeuAlaTyr--------- 370

Qy    769 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 828
Db    371 -------------------------------------------------------Asp 371

Qy    829 ACGGATCGCTAC------------------------------------CTCACTTCTTGG 852
Db    372 LysAspArgTyrThrGluGluGluArgGluValTyrSerTyrLeuArgTyrThrGlyThr 391

Qy    853 GGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATC------------------ 894
Db    392 ProIleProAspAsnProAsnAspLysAsnAsnSerGlnLeuValValSerValAlaGly 411

Qy    895 GAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACC 954
Db    412 ThrValGluGlyThrAsnGlnAspIleSerLeuLysPhePheGluIleAspLeuThrSer 431

Qy    955 CGTATCAAATATGCCAGCAAGTATGCTCAG---------------AGTGCTAACTATTGG 999
Db    432 ArgProAlaHisGlyGlyLysThrGluGlnGlyLeuSerProLysSerLysProPheAla 451

Qy   1000 AAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGT 1059
Db    452 ThrAspSerGlyAlaMetSerHisLysLeuGluLysAlaAspLeuLeu------LysAla 469

Qy   1060 GCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAAC--------------------- 1098
Db    470 IleGlnGluGlnLeuIleAlaAsn---ValHisSerAsnAspAspTyrPheGluValIle 488

Qy   1099 ------GGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAG 1152
Db    489 AspPheAlaSerAspAlaThrIleThrAspArgAsnGlyLysValTyrPheAlaAspLys 508

Qy   1153 GAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACC 1212
Db    509 AspGlySer---ValThrLeuProThrGlnProValGlnGluPheLeuLeuSerGlyHis 527

Qy   1213 GAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCC 1272
Db    528 ValArgValArgTyrLysGlu--------------------LysProIleGlnAsnGln 540

Qy   1273 GGTATCCTCAAATCGCTTGACGACAAGTAC----------------------------- 1302
Db    541 Ala------LysSerValAspValGluTyrThrValGlnPheThrProLeuAsnProAsp 558

Qy   1303 AAAGACTACCTCCCCTCGCTC---GACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTA 1359
Db    559 AspAspPheArgProGlyLeuLysAspThrLysLeuLeuLysThrLeu------AlaIle 576

Qy   1360 CGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAA 1419
Db    577 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 596

Qy   1420 TTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTAT 1479
Db    597 HisProGly-----------TyrThrIleTyrGluArgAspSerSerIleValThrHis 612

Qy   1480 AGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATC 1539
Db    613 AspAsnAspIlePheArgThrIleLeuProMetAspGlnGluPheThrTyrArgValLys 632

Qy   1540 GAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAG 1599
Db    633 AsnArgGluGlnAlaTyrArgIleAsnLysLysSerGlyLeuAsnGluGluIleAsnAsn 652

Qy   1600 GCCGATGCGATGGCCAATGCC---TATGCCATTGAGAAGGGCAAG-------------- 1641
```

```
Db    653 ThrAspLeuIleSerGluLysTyrTyrValLeuLysLysGlyGluLysProTyrAspPro 672

Qy   1642 -----------------CGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGT 1683
                           :::||||||               ::::::
Db    673 PheAspArgSerHisLeuLysLeuPheThrIleLysTyrValAspValAspThrAsnGlu 692

Qy   1684 GCTCTGCCGAGCGATGCCAACTTCACC---------ATGCGTATGAGCTACGGCTCCATC 1734
             |||   |||::.             |||           :::   :::      :::
Db    693 LeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArgAsnLeuAspPheArgAspLeu 712

Qy   1735 AAGGGATATGAACCGCAGGACGGTGCC------TGGTACAAC------------------ 1770
                ||| ::: |||:::|||   |||         ||||||
Db    713 ------TyrAspProArgAspLysAlaLysLeuLeuTyrAsnAsnLeuAspAlaPheGly 730

Qy   1771 ---------TATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGAT 1821
                   |||      |||||||||
Db    731 IleMetAspTyrThrLeuThrGlyLys------------------------------ 739

Qy   1822 GAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC------ACCAAAAACTATGGT 1875
                   |||::::::|||        |||       |||       |||        |||
Db    740 ---------ValGluAspAsnHisAspAspThrAsnArgIleIleThrValTyrMetGly 756

Qy   1876 CGC------TATGCCGAGAACGGTCAGCTCCATATCGCTTTC 1911
           :::             ||||||         |||:::|||:::
Db    757 LysArgProGluGlyGluAsnAlaSerTyrHisLeuAlaTyr 770
```

RESULT 7
US-08-488-940-3
; Sequence 3, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,940
; FILING DATE: 09-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 813 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-488-940-3

Alignment Scores:

Pred. No.:	0.00438	Length:	813
Score:	122.50	Matches:	130
Percent Similarity:	33.98%	Conservative:	99
Best Local Similarity:	19.29%	Mismatches:	254
Query Match:	3.21%	Indels:	191
DB:	2	Gaps:	35

US-10-008-355-1 (1-2139) x US-08-488-940-3 (1-813)

```
Qy    223 GGTATCACAGTGTCCGATCAGGGCCTG------------------ATCTTTACCAACCAC 264
          |||   |||::::::   |||   |||                     |||       :::
Db    190 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 209

Qy    265 CACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGAT 324
          |||   ||||||          :::          |||::::::                |||
Db    210 HisProGlyTyrThrIleTyrGluArgAspSerSerIle---------ValThrHisAsp 226

Qy    325 GGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCG---------GGTCTTTCC 375
             :::   ||||||:::             ||||||:::
Db    227 AsnAspIlePheArgThrIle---------LeuProMetAspGlnGluPheThrTyrArg 243

Qy    376 GTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGT 435
          ||||||        :::      ::::::       |||                   |||
Db    244 ValLysAsnArgGluGlnAlaTyrArgIleAsnLysLys---------------SerGly 258

Qy    436 ATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAA 495
          :::   :::|||:::                        |||          |||   |||
Db    259 LeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLysLys 278

Qy    496 AAAGAA------------------------------------------------------ 501
             |||
Db    279 GlyGluLysProTyrAspProPheAspArgSerHisLeuLysLeuPheThrIleLysTyr 298

Qy    502 ---AATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATAC 558
             :::   |||   |||:::|||          |||          :::    :::|||
Db    299 ValAspValAspThrAsnGluLeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArg 318

Qy    559 ---------TTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCT 609
                   |||      :::||||||                ::::::::::::
Db    319 AsnLeuAspPheArgAspLeuTyrAspProArgAspLysAlaLysLeuLeuTyr------ 336

Qy    610 CCCAGCTCTGTAGGTAAGTTCGGA--------------------GGCGATACGGACAAC 648
             :::::::::      ||||||                      |||      ::::::
Db    337 ---AsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLysValGluAsp 355

Qy    649 TGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGAC 708
                         |||          :::    ||||||   |||
Db    356 -----------AsnHisAspAspThrAsnArgIleIleThrValTyrMetGly------ 369

Qy    709 AACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCC 768
             ||||||                :::|||      |||           |||
Db    370 LysArgPro---------GluGlyGluAsnAlaSerTyrHisLeuAlaTyr--------- 383

Qy    769 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 828

Db    384 ---------------------------------------------------------Asp 384

Qy    829 ACGGATCGCTAC------------------------------------CTCACTTCTTGG 852
             |||||||||                                             |||
Db    385 LysAspArgTyrThrGluGluGluArgGluValTyrSerTyrLeuArgTyrThrGlyThr 404

Qy    853 GGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATC------------------ 894
             :::   |||             ::::::||||||   ::::::
Db    405 ProIleProAspAsnProAsnAspLysAsnAsnSerGlnLeuValValSerValAlaGly 424

Qy    895 GAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACC 954
             |||   |||       |||   |||        :::                |||   :::
Db    425 ThrValGluGlyThrAsnGlnAspIleSerLeuLysPhePheGluIleAspLeuThrSer 444

Qy    955 CGTATCAAATATGCCAGCAAGTATGCTCAG---------------AGTGCTAACTATTGG 999
          |||      :::      |||      |||                   |||      :::
Db    445 ArgProAlaHisGlyGlyLysThrGluGlnGlyLeuSerProLysSerLysProPheAla 464

Qy   1000 AAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGT 1059
             :::|||       |||:::          |||      :::   |||::::::     |||
Db    465 ThrAspSerGlyAlaMetSerHisLysLeuGluLysAlaAspLeuLeu------LysAla 482

Qy   1060 GCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAAC--------------------- 1098
```

```
                 :::|||:::      |||:::    :::        |||
Db    483 IleGlnGluGlnLeuIleAlaAsn---ValHisSerAsnAspAspTyrPheGluValIle 501

Qy   1099 ------GGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAG 1152
                          |||        |||              :::       |||   |||
Db    502 AspPheAlaSerAspAlaThrIleThrAspArgAsnGlyLysValTyrPheAlaAspLys 521

Qy   1153 GAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACC 1212
          :::|||:::                       :::   |||    |||       |||
Db    522 AspGlySer---ValThrLeuProThrGlnProValGlnGluPheLeuLeuSerGlyHis 540

Qy   1213 GAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCC 1272
             ||||||:::   :::                                  |||     :::
Db    541 ValArgValArgTyrLysGlu--------------------LysProIleGlnAsnGln 553

Qy   1273 GGTATCCTCAAATCGCTTGACGACAAGTAC------------------------------ 1302
                   ||||||:::|||   :::|||
Db    554 Ala------LysSerValAspValGluTyrThrValGlnPheThrProLeuAsnProAsp 571

Qy   1303 AAAGACTACCTCCCCTCGCTC---GACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTA 1359
             |||:::   |||   |||   |||   |||:::|||          :::        :::
Db    572 AspAspPheArgProGlyLeuLysAspThrLysLeuLeuLysThrLeu------AlaIle 589

Qy   1360 CGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAA 1419
                |||   :::   :::|||             ::::::::::::::::|||
Db    590 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 609

Qy   1420 TTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTAT 1479
                |||             |||        :::      |||    |||:::|||    :::
Db    610 HisProGly-----------TyrThrIleTyrGluArgAspSerSerIleValThrHis 625

Qy   1480 AGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATC 1539
             :::                :::      ||||||:::|||          :::
Db    626 AspAsnAspIlePheArgThrIleLeuProMetAspGlnGluPheThrTyrArgValLys 645

Qy   1540 GAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAG 1599
             ::::::   |||      ::::::|||
Db    646 AsnArgGluGlnAlaTyrArgIleAsnLysLysSerGlyLeuAsnGluGluIleAsnAsn 665

Qy   1600 GCCGATGCGATGGCCAATGCC---TATGCCATTGAGAAGGGCAAG--------------- 1641
                |||   ::::::      |||    ::::::||||||:::
Db    666 ThrAspLeuIleSerGluLysTyrTyrValLeuLysLysGlyGluLysProTyrAspPro 685

Qy   1642 ------------------CGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGT 1683
                             :::||||||              ::::::
Db    686 PheAspArgSerHisLeuLysLeuPheThrIleLysTyrValAspValAspThrAsnGlu 705

Qy   1684 GCTCTGCCGAGCGATGCCAACTTCACC---------ATGCGTATGAGCTACGGCTCCATC 1734
             |||   |||:::         |||              :::   :::        :::
Db    706 LeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArgAsnLeuAspPheArgAspLeu 725

Qy   1735 AAGGGATATGAACCGCAGGACGGTGCC------TGGTACAAC------------------ 1770
                |||:::|||:::|||   |||         ||||||
Db    726 ------TyrAspProArgAspLysAlaLysLeuLeuTyrAsnAsnLeuAspAlaPheGly 743

Qy   1771 ---------TATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGAT 1821
                   |||      |||||||||
Db    744 IleMetAspTyrThrLeuThrGlyLys-------------------------------- 752

Qy   1822 GAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC------ACCAAAAACTATGGT 1875
                   |||::::::|||      |||      |||         |||         |||
Db    753 ---------ValGluAspAsnHisAspAspThrAsnArgIleIleThrValTyrMetGly 769

Qy   1876 CGC------TATGCCGAGAACGGTCAGCTCCATATCGCTTTC 1911
          :::                ||||||      |||::::||||:::
Db    770 LysArgProGluGlyGluAsnAlaSerTyrHisLeuAlaTyr 783
```

RESULT 8
US-08-488-940-2
; Sequence 2, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,940
; FILING DATE: 09-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1181 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-488-940-2

```
Alignment Scores:
Pred. No.:                0.00589       Length:          1181
Score:                    122.50        Matches:         130
Percent Similarity:       33.98%        Conservative:    99
Best Local Similarity:    19.29%        Mismatches:      254
Query Match:              3.21%         Indels:          191
DB:                       2             Gaps:            35
```

US-10-008-355-1 (1-2139) x US-08-488-940-2 (1-1181)

```
Qy    223 GGTATCACAGTGTCCGATCAGGGCCTG-----------------ATCTTTACCAACCAC 264
          |||   |||::::::   |||   |||                   |||         :::
Db    558 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 577

Qy    265 CACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGAT 324
          |||   ||||||            :::      |||::::::                |||
Db    578 HisProGlyTyrThrIleTyrGluArgAspSerSerIle---------ValThrHisAsp 594

Qy    325 GGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCG---------GGTCTTTCC 375
                :::   ||||||:::         ||||||:::
Db    595 AsnAspIlePheArgThrIle---------LeuProMetAspGlnGluPheThrTyrArg 611

Qy    376 GTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGT 435
          ||||||         :::      ::::::         |||                |||
Db    612 ValLysAsnArgGluGlnAlaTyrArgIleAsnLysLys---------------SerGly 626

Qy    436 ATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAA 495
          :::      :::|||:::                        |||         |||   |||
Db    627 LeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLysLys 646

Qy    496 AAAGAA------------------------------------------------------ 501
             |||
Db    647 GlyGluLysProTyrAspProPheAspArgSerHisLeuLysLeuPheThrIleLysTyr 666

Qy    502 ---AATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATAC 558
             :::   |||   |||:::|||         |||         :::   :::|||
Db    667 ValAspValAspThrAsnGluLeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArg 686
```

```
Qy     559 ---------TTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCT 609
                     |||     :::||||||             :::::::::::
Db     687 AsnLeuAspPheArgAspLeuTyrAspProArgAspLysAlaLysLeuLeuTyr------ 704

Qy     610 CCCAGCTCTGTAGGTAAGTTCGGA---------------------GGCGATACGGACAAC 648
              ::::::::     ||||||                       |||      ::::::
Db     705 ---AsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLysValGluAsp 723

Qy     649 TGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGAC 708
                          |||          :::      ||||||   |||
Db     724 -----------AsnHisAspAspThrAsnArgIleIleThrValTyrMetGly------ 737

Qy     709 AACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCC 768
              ||||||             :::|||       |||       |||
Db     738 LysArgPro---------GluGlyGluAsnAlaSerTyrHisLeuAlaTyr--------- 751

Qy     769 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 828

Db     752 ---------------------------------------------------------Asp 752

Qy     829 ACGGATCGCTAC------------------------------------CTCACTTCTTGG 852
              |||||||||                                     |||
Db     753 LysAspArgTyrThrGluGluGluArgGluValTyrSerTyrLeuArgTyrThrGlyThr 772

Qy     853 GGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATC------------------ 894
              :::   |||        ::::::||||||   ::::::
Db     773 ProIleProAspAsnProAsnAspLysAsnAsnSerGlnLeuValValSerValAlaGly 792

Qy     895 GAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACC 954
              |||   |||      |||   |||      :::             |||       :::
Db     793 ThrValGluGlyThrAsnGlnAspIleSerLeuLysPhePheGluIleAspLeuThrSer 812

Qy     955 CGTATCAAATATGCCAGCAAGTATGCTCAG---------------AGTGCTAACTATTGG 999
              |||      :::      |||      |||               |||       :::
Db     813 ArgProAlaHisGlyGlyLysThrGluGlnGlyLeuSerProLysSerLysProPheAla 832

Qy    1000 AAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGT 1059
              :::|||      |||:::      |||   :::   |||::::::        |||
Db     833 ThrAspSerGlyAlaMetSerHisLysLeuGluLysAlaAspLeuLeu------LysAla 850

Qy    1060 GCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAAC--------------------- 1098
              :::|||:::      |||:::   :::      |||
Db     851 IleGlnGluGlnLeuIleAlaAsn---ValHisSerAsnAspAspTyrPheGluValIle 869

Qy    1099 ------GGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAG 1152
                          |||      |||         :::      |||   |||
Db     870 AspPheAlaSerAspAlaThrIleThrAspArgAsnGlyLysValTyrPheAlaAspLys 889

Qy    1153 GAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACC 1212
              :::|||:::                     :::   |||   |||      |||
Db     890 AspGlySer---ValThrLeuProThrGlnProValGlnGluPheLeuLeuSerGlyHis 908

Qy    1213 GAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCC 1272
              ||||||:::   :::                           |||       :::
Db     909 ValArgValArgTyrLysGlu--------------------LysProIleGlnAsnGln 921

Qy    1273 GGTATCCTCAAATCGCTTGACGACAAGTAC------------------------------ 1302
                    ||||||:::|||   :::|||
Db     922 Ala------LysSerValAspValGluTyrThrValGlnPheThrProLeuAsnProAsp 939

Qy    1303 AAAGACTACCTCCCCTCGCTC---GACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTA 1359
              |||:::   |||   |||   |||   |||:::|||          :::       :::
Db     940 AspAspPheArgProGlyLeuLysAspThrLysLeuLeuLysThrLeu------AlaIle 957

Qy    1360 CGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAA 1419
              |||   :::   :::|||             :::::::::::::::|||
Db     958 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 977

Qy    1420 TTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTAT 1479
                    |||             |||   :::      |||   |||:::|||       :::
Db     978 HisProGly------------TyrThrIleTyrGluArgAspSerSerIleValThrHis 993

Qy    1480 AGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATC 1539
              :::            :::      ||||||:::|||          :::
Db     994 AspAsnAspIlePheArgThrIleLeuProMetAspGlnGluPheThrTyrArgValLys 1013

Qy    1540 GAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAG 1599
              ::::::   |||      :::::::|||
Db    1014 AsnArgGluGlnAlaTyrArgIleAsnLysLysSerGlyLeuAsnGluGluIleAsnAsn 1033

Qy    1600 GCCGATGCGATGGCCAATGCC---TATGCCATTGAGAAGGGCAAG--------------- 1641
              |||   ::::::            |||   ::::::||||||:::
Db    1034 ThrAspLeuIleSerGluLysTyrTyrValLeuLysLysGlyGluLysProTyrAspPro 1053

Qy    1642 ------------------CGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGT 1683
                              :::||||||            ::::::
Db    1054 PheAspArgSerHisLeuLysLeuPheThrIleLysTyrValAspValAspThrAsnGlu 1073

Qy    1684 GCTCTGCCGAGCGATGCCAACTTCACC---------ATGCGTATGAGCTACGGCTCCATC 1734
              |||   |||:::            |||                   :::   :::   :::
Db    1074 LeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArgAsnLeuAspPheArgAspLeu 1093

Qy    1735 AAGGGATATGAACCGCAGGACGGTGCC------TGGTACAAC------------------ 1770
                 |||:::|||:::|||   |||         ||||||
Db    1094 ------TyrAspProArgAspLysAlaLysLeuLeuTyrAsnAsnLeuAspAlaPheGly 1111

Qy    1771 ---------TATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGAT 1821
                    |||      |||||||||
Db    1112 IleMetAspTyrThrLeuThrGlyLys--------------------------------- 1120

Qy    1822 GAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC------ACCAAAAACTATGGT 1875
                    |||::::::|||      |||      |||      |||         |||
Db    1121 ---------ValGluAspAsnHisAspAspThrAsnArgIleIleThrValTyrMetGly 1137

Qy    1876 CGC------TATGCCGAGAACGGTCAGCTCCATATCGCTTTC 1911
              :::            ||||||      |||:::|||:::
Db    1138 LysArgProGluGlyGluAsnAlaSerTyrHisLeuAlaTyr 1151

RESULT 9
US-08-488-940-1
; Sequence 1, Application US/08488940
; Patent No. 5854049
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Guy L.
;    TITLE OF INVENTION:  PLASMIN-RESISTANT STREPTOKINASE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,940
;      FILING DATE:  09-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  05433/009001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1194 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
US-08-488-940-1

Alignment Scores:
Pred. No.:                0.00594       Length:        1194
Score:                    122.50        Matches:       130
Percent Similarity:       33.98%        Conservative:  99
Best Local Similarity:    19.29%        Mismatches:    254
Query Match:              3.21%         Indels:        191
DB:                       2             Gaps:          35

US-10-008-355-1 (1-2139) x US-08-488-940-1 (1-1194)

Qy     223 GGTATCACAGTGTCCGATCAGGGCCTG-----------------ATCTTTACCAACCAC 264
Db     571 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 590

Qy     265 CACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGAT 324
Db     591 HisProGlyTyrThrIleTyrGluArgAspSerSerIle---------ValThrHisAsp 607

Qy     325 GGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCG---------GGTCTTTCC 375
Db     608 AsnAspIlePheArgThrIle---------LeuProMetAspGlnGluPheThrTyrArg 624

Qy     376 GTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGT 435
Db     625 ValLysAsnArgGluGlnAlaTyrArgIleAsnLysLys---------------SerGly 639

Qy     436 ATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAA 495
Db     640 LeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLysLys 659

Qy     496 AAAGAA------------------------------------------------------ 501
Db     660 GlyGluLysProTyrAspProPheAspArgSerHisLeuLysLeuPheThrIleLysTyr 679

Qy     502 ---AATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATAC 558
Db     680 ValAspValAspThrAsnGluLeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArg 699

Qy     559 ---------TTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCT 609
Db     700 AsnLeuAspPheArgAspLeuTyrAspProArgAspLysAlaLysLeuLeuTyr------ 717

Qy     610 CCCAGCTCTGTAGGTAAGTTCGGA--------------------GGCGATACGGACAAC 648
Db     718 ---AsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLysValGluAsp 736

Qy     649 TGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGAC 708
Db     737 -----------AsnHisAspAspThrAsnArgIleIleThrValTyrMetGly------ 750

Qy     709 AACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCC 768
Db     751 LysArgPro---------GluGlyGluAsnAlaSerTyrHisLeuAlaTyr--------- 764

Qy     769 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 828
Db     765 ---------------------------------------------------------Asp 765

Qy     829 ACGGATCGCTAC------------------------------------CTCACTTCTTGG 852
Db     766 LysAspArgTyrThrGluGluGluArgGluValTyrSerTyrLeuArgTyrThrGlyThr 785

Qy     853 GGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATC------------------ 894
Db     786 ProIleProAspAsnProAsnAspLysAsnAsnSerGlnLeuValValSerValAlaGly 805

Qy     895 GAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACC 954
Db     806 ThrValGluGlyThrAsnGlnAspIleSerLeuLysPhePheGluIleAspLeuThrSer 825

Qy     955 CGTATCAAATATGCCAGCAAGTATGCTCAG---------------AGTGCTAACTATTGG 999
Db     826 ArgProAlaHisGlyGlyLysThrGluGlnGlyLeuSerProLysSerLysProPheAla 845

Qy    1000 AAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGT 1059
Db     846 ThrAspSerGlyAlaMetSerHisLysLeuGluLysAlaAspLeuLeu------LysAla 863

Qy    1060 GCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAAC-------------------- 1098
Db     864 IleGlnGluGlnLeuIleAlaAsn---ValHisSerAsnAspAspTyrPheGluValIle 882

Qy    1099 ------GGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAG 1152
Db     883 AspPheAlaSerAspAlaThrIleThrAspArgAsnGlyLysValTyrPheAlaAspLys 902

Qy    1153 GAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACC 1212
Db     903 AspGlySer---ValThrLeuProThrGlnProValGlnGluPheLeuLeuSerGlyHis 921

Qy    1213 GAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCC 1272
Db     922 ValArgValArgTyrLysGlu--------------------LysProIleGlnAsnGln 934

Qy    1273 GGTATCCTCAAATCGCTTGACGACAAGTAC----------------------------- 1302
Db     935 Ala------LysSerValAspValGluTyrThrValGlnPheThrProLeuAsnProAsp 952

Qy    1303 AAAGACTACCTCCCCTCGCTC---GACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTA 1359
Db     953 AspAspPheArgProGlyLeuLysAspThrLysLeuLeuLysThrLeu------AlaIle 970

Qy    1360 CGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAA 1419
Db     971 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 990

Qy    1420 TTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTAT 1479
Db     991 HisProGly------------TyrThrIleTyrGluArgAspSerSerIleValThrHis 1006

Qy    1480 AGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATC 1539
Db    1007 AspAsnAspIlePheArgThrIleLeuProMetAspGlnGluPheThrTyrArgValLys 1026

Qy    1540 GAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAG 1599
Db    1027 AsnArgGluGlnAlaTyrArgIleAsnLysLysSerGlyLeuAsnGluGluIleAsnAsn 1046

Qy    1600 GCCGATGCGATGGCCAATGCC---TATGCCATTGAGAAGGGCAAG--------------- 1641
Db    1047 ThrAspLeuIleSerGluLysTyrTyrValLeuLysLysGlyGluLysProTyrAspPro 1066

Qy    1642 ------------------CGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGT 1683
Db    1067 PheAspArgSerHisLeuLysLeuPheThrIleLysTyrValAspValAspThrAsnGlu 1086

Qy    1684 GCTCTGCCGAGCGATGCCAACTTCACC---------ATGCGTATGAGCTACGGCTCCATC 1734
Db    1087 LeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArgAsnLeuAspPheArgAspLeu 1106

Qy    1735 AAGGGATATGAACCGCAGGACGGTGCC------TGGTACAAC------------------ 1770
Db    1107 ------TyrAspProArgAspLysAlaLysLeuLeuTyrAsnAsnLeuAspAlaPheGly 1124

Qy    1771 ---------TATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGAT 1821
Db    1125 IleMetAspTyrThrLeuThrGlyLys--------------------------------- 1133

Qy    1822 GAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC------ACCAAAAACTATGGT 1875
Db    1134 ---------ValGluAspAsnHisAspAspThrAsnArgIleIleThrValTyrMetGly 1150

Qy    1876 CGC------TATGCCGAGAACGGTCAGCTCCATATCGCTTTC 1911
```

```
             :::           ||||||         |||:::|||:::
Db    1151 LysArgProGluGlyGluAsnAlaSerTyrHisLeuAlaTyr 1164

RESULT 10
US-08-488-940-17
; Sequence 17, Application US/08488940
; Patent No. 5854049
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Guy L.
;    TITLE OF INVENTION:  PLASMIN-RESISTANT STREPTOKINASE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,940
;      FILING DATE:  09-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  05433/009001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1194 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-488-940-17

Alignment Scores:
Pred. No.:              0.00594      Length:        1194
Score:                  122.50       Matches:       130
Percent Similarity:     33.98%       Conservative:  99
Best Local Similarity:  19.29%       Mismatches:    254
Query Match:            3.21%        Indels:        191
DB:                     2            Gaps:          35

US-10-008-355-1 (1-2139) x US-08-488-940-17 (1-1194)

Qy     223 GGTATCACAGTGTCCGATCAGGGCCTG-----------------ATCTTTACCAACCAC 264
           |||   |||::::::  |||   |||                 |||        :::
Db     571 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 590

Qy     265 CACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGAT 324
           |||   ||||||         :::      |||::::::                  |||
Db     591 HisProGlyTyrThrIleTyrGluArgAspSerSerIle---------ValThrHisAsp 607

Qy     325 GGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCG---------GGTCTTTCC 375
              :::   ||||||:::            ||||||:::
Db     608 AsnAspIlePheArgThrIle---------LeuProMetAspGlnGluPheThrTyrArg 624

Qy     376 GTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGT 435
           ||||||         :::      ::::::      |||                  |||
Db     625 ValLysAsnArgGluGlnAlaTyrArgIleAsnLysLys---------------SerGly 639

Qy     436 ATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAA 495
           :::   :::|||:::                        |||         |||   |||
Db     640 LeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLysLys 659

Qy     496 AAAGAA------------------------------------------------------ 501
              |||
Db     660 GlyGluLysProTyrAspProPheAspArgSerHisLeuLysLeuPheThrIleLysTyr 679

Qy     502 ---AATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATAC 558
              :::   |||   |||:::|||            |||         :::   :::|||
Db     680 ValAspValAspThrAsnGluLeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArg 699

Qy     559 ---------TTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCT 609
                    |||      :::||||||            ::::::::::::
Db     700 AsnLeuAspPheArgAspLeuTyrAspProArgAspLysAlaLysLeuLeuTyr------ 717

Qy     610 CCCAGCTCTGTAGGTAAGTTCGGA--------------------GGCGATACGGACAAC 648
              :::::::::         ||||||                  |||         ::::::
Db     718 ---AsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLysValGluAsp 736

Qy     649 TGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGAC 708
                       |||            :::      ||||||   |||
Db     737 -----------AsnHisAspAspThrAsnArgIleIleThrValTyrMetGly------ 750

Qy     709 AACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCC 768
              ||||||            :::|||      |||         |||
Db     751 LysArgPro---------GluGlyGluAsnAlaSerTyrHisLeuAlaTyr--------- 764

Qy     769 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 828

Db     765 ---------------------------------------------------------Asp 765

Qy     829 ACGGATCGCTAC-----------------------------------CTCACTTCTTGG 852
              |||||||||                                       |||
Db     766 LysAspArgTyrThrGluGluGluArgGluValTyrSerTyrLeuArgTyrThrGlyThr 785

Qy     853 GGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATC------------------ 894
              :::   |||         ::::::||||||   ::::::
Db     786 ProIleProAspAsnProAsnAspLysAsnAsnSerGlnLeuValValSerValAlaGly 805

Qy     895 GAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACC 954
              |||   |||      |||   |||      :::         |||      :::
Db     806 ThrValGluGlyThrAsnGlnAspIleSerLeuLysPhePheGluIleAspLeuThrSer 825

Qy     955 CGTATCAAATATGCCAGCAAGTATGCTCAG---------------AGTGCTAACTATTGG 999
           |||      :::      |||      |||               |||      :::
Db     826 ArgProAlaHisGlyGlyLysThrGluGlnGlyLeuSerProLysSerLysProPheAla 845

Qy    1000 AAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGT 1059
           :::|||      |||:::         |||      :::   |||::::::      |||
Db     846 ThrAspSerGlyAlaMetSerHisLysLeuGluLysAlaAspLeuLeu------LysAla 863

Qy    1060 GCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAAC--------------------- 1098
           :::|||:::         |||:::      :::      |||
Db     864 IleGlnGluGlnLeuIleAlaAsn---ValHisSerAsnAspAspTyrPheGluValIle 882

Qy    1099 ------GGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAG 1152
                    |||         |||                  :::      |||   |||
Db     883 AspPheAlaSerAspAlaThrIleThrAspArgAsnGlyLysValTyrPheAlaAspLys 902

Qy    1153 GAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACC 1212
           :::|||:::                              :::   |||   |||      |||
Db     903 AspGlySer---ValThrLeuProThrGlnProValGlnGluPheLeuLeuSerGlyHis 921

Qy    1213 GAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCC 1272
              ||||||:::   :::                              |||      :::
Db     922 ValArgValArgTyrLysGlu--------------------LysProIleGlnAsnGln 934

Qy    1273 GGTATCCTCAAATCGCTTGACGACAAGTAC------------------------------ 1302
                    ||||||:::|||   :::|||
Db     935 Ala------LysSerValAspValGluTyrThrValGlnPheThrProLeuAsnProAsp 952

Qy    1303 AAAGACTACCTCCCCTCGCTC---GACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTA 1359
              |||:::   |||   |||   |||   |||:::|||            :::      :::
Db     953 AspAspPheArgProGlyLeuLysAspThrLysLeuLeuLysThrLeu------AlaIle 970
```

```
Qy   1360 CGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAA 1419
                  |||   :::   :::|||              ::::::::::::::|||
Db    971 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 990

Qy   1420 TTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTAT 1479
                  |||     .     |||     :::     |||   |||:::|||   :::
Db    991 HisProGly------------TyrThrIleTyrGluArgAspSerSerIleValThrHis 1006

Qy   1480 AGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATC 1539
             :::               :::     ||||||:::|||         :::
Db   1007 AspAsnAspIlePheArgThrIleLeuProMetAspGlnGluPheThrTyrArgValLys 1026

Qy   1540 GAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAG 1599
             ::::::   |||       ::::::|||
Db   1027 AsnArgGluGlnAlaTyrArgIleAsnLysLysSerGlyLeuAsnGluGluIleAsnAsn 1046

Qy   1600 GCCGATGCGATGGCCAATGCC---TATGCCATTGAGAAGGGCAAG--------------- 1641
             |||   ::::::         |||   ::::::||||||:::
Db   1047 ThrAspLeuIleSerGluLysTyrTyrValLeuLysLysGlyGluLysProTyrAspPro 1066

Qy   1642 -----------------CGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGT 1683
                            :::||||||               ::::::
Db   1067 PheAspArgSerHisLeuLysLeuPheThrIleLysTyrValAspValAspThrAsnGlu 1086

Qy   1684 GCTCTGCCGAGCGATGCCAACTTCACC---------ATGCGTATGAGCTACGGCTCCATC 1734
             |||   |||:::         |||               :::   :::      :::
Db   1087 LeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArgAsnLeuAspPheArgAspLeu 1106

Qy   1735 AAGGGATATGAACCGCAGGACGGTGCC------TGGTACAAC------------------ 1770
                |||:::|||:::|||.  |||         ||||||
Db   1107 ------TyrAspProArgAspLysAlaLysLeuLeuTyrAsnAsnLeuAspAlaPheGly 1124

Qy   1771 ---------TATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGAT 1821
                   |||      |||||||||
Db   1125 IleMetAspTyrThrLeuThrGlyLys--------------------------------- 1133

Qy   1822 GAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC------ACCAAAAACTATGGT 1875
                   |||:::::|||      |||      |||      |||         |||
Db   1134 ---------ValGluAspAsnHisAspAspThrAsnArgIleIleThrValTyrMetGly 1150

Qy   1876 CGC------TATGCCGAGAACGGTCAGCTCCATATCGCTTTC 1911
             :::            ||||||         |||:::|||:::
Db   1151 LysArgProGluGlyGluAsnAlaSerTyrHisLeuAlaTyr 1164
```

RESULT 11
US-09-036-987A-3
; Sequence 3, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2152 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-036-987A-3

```
Alignment Scores:
Pred. No.:              0.0213      Length:        2152
Score:                  119.00      Matches:       161
Percent Similarity:     30.96%      Conservative:  100
Best Local Similarity:  19.10%      Mismatches:    274
Query Match:            3.01%       Indels:        308
DB:                     4           Gaps:          33
```

US-10-008-355-1 (1-2139) x US-09-036-987A-3 (1-2152)

```
Qy   2139 TTAGATCAACTTCAGCTCTTGGATGAGACGGGGGCACTGACCCCATTTGTCAATCATGAA 2080
             ||||||   |||   |||:::||||||         |||      |||      :::   |||
Db   1415 LeuAspGlyLeuAlaLeuValAspGluProThrAlaThrAlaProLeuGlyAspGlyGlu 1434

Qy   2079 GAGAACGTAGCGGATGTCCACGCT--------------GATTGTGCGCTGCAGATCGGG 2035
                :::|||      |||               |||      :::   :::|||
Db   1435 ValArgIleAlaMetArgAlaAlaGlyValAsnPheArgAspAlaLeuIleAlaLeuGly 1454

Qy   2034 TTCGAACTCGATGTCACCACTCATAGCTTCCCAGTTGCCATCGAAAGCAAGACCGATCAG 1975
                                 :::|||    .          ::::::
Db   1455 Met-----------------------TyrProGlyValAlaSerLeuGlySerGluGly 1466

Qy   1974 ACGGCCGTTCTTATCGAATAC------GGGGCTACCGGAGTTACCGCCCGTGATGTCGTT 1921
                |||::::::|||         |||  .  ||||||:::         |||
Db   1467 AlaGlyValValValGluThrGlyProGlyValThrGlyLeuAlaProGlyAsp------ 1484

Qy   1920 GTTCGATAGGAAAGCGATATGGAGCTGACCGTTCTCGGCATAGCGACCATAGTTTTTGGT 1861
                                          |||:::|||:::         ||||||
Db   1485 ---------------------------ArgValMetGlyMetIleProLysAlaPheGly 1495

Qy   1860 GCGGAAGAGGTCGAGGATATTCTCCTGTACGGCAAACTC-------------------- 1822
                |||   |||      :::      ::::::
Db   1496 ProLeuAlaValAlaAspHisArgMetValThrArgIleProAlaGlyTrpSerPheAla 1515

Qy   1821 ------------ATCGCTCTTAGGATCCTGCTTCTCCAATACGCCCTTGCCTGTCGTATG 1774
                         :::   :::      |||         |||||||||
Db   1516 ArgAlaAlaSerValProIleValPheLeuThrAlaTyrTyrAlaLeuValAspLeuAla 1535

Qy   1773 ATAGTTGTACCAGGCACCGTCCTGCGGTTCATATCCCTTGATGGAGCCGTAGCTCATACG 1714
                :::   ||||||      |||      |||      |||   |||
Db   1536 GlyLeuArgProGlyGluSerLeuLeuValHisSerAlaAlaGlyGlyValGly------ 1553

Qy   1713 CATGGTGAAGTTGGCATCGCTCGGCAGAGCACGTCCGGGGTACATCTCACGCAAACCGGC 1654
                   :::   :::||||||      :::   |||:::   |||   :::
Db   1554 MetAlaAlaIleGlnLeuAlaArgHisLeuGlyAlaGluValTyrAlaThrAlaSerGlu 1573

Qy   1654 ------------------------------------------------------------ 1654

Db   1574 AspLysTrpGlnAlaValGluLeuSerArgGluHisLeuAlaSerSerArgThrCysAsp 1593

Qy   1653 ---AAAGAAAAGACGCTTGCCCTTCTCAATGGCATAGGCATTGGCCATCGCATCGGCCTG 1597
             ::::::::   |||         |||   |||:::         :::   |||
Db   1594 PheGluGlnGlnPheLeuGlyAlaThrGlyGlyArgGlyValAsp------ValValLeu 1611

Qy   1596 AATAGCGCGAGCAGCAGCTATTACGCTCTTGGAAAGCTCTACTGCCGGATCTTTCTCGAT 1537
```

```
               ||||||   :::                                      |||   |||
Db    1612 AsnSerLeuAlaGlyGlu---------------------------------PheAlaAsp 1620

Qy    1536 AGC------------------CTTGGCAAACTTTTCCTTGTCCATGGA------------ 1507
           :::                     |||      ||||||      |||
Db    1621 AlaSerLeuArgMetLeuProArgGlyGlyArgPheLeuGluLeuGlyLysThrAspVal 1640

Qy    1507 ------------------------------------------------------------ 1507

Db    1641 ArgAspProValGluValAlaAspAlaHisProGlyValSerTyrGlnAlaPheAspThr 1660

Qy    1506 CTTGAGCATGGCATGGAACTTGTCGCTATAAGGAACCACACTCTTGTCGAA---TACGAA 1450
           :::|||   |||   :::   :::               |||   |||||||||   :::|||
Db    1661 ValGluAlaGlyProGlnArgIleGlyGluMetLeuHisGluLeuValGluLeuPheGlu 1680

Qy    1449 GTCTGCATACTT---------------------------CTTCGTGTCGCCTTTGAATTT 1417
                 :::|||                            :::|||   |||   |||
Db    1681 GlyArgValLeuGluProLeuProValThrAlaTrpAspValArgGlnAlaProGluAla 1700

Qy    1416 CTTGTCGATTACATT---CTTGAATATATCGGGGAGCTTGTCGGCAGGGATACGCCGGCG 1360
           |||                     ::::::|||:::||||||            |||
Db    1701 LeuArgHisLeuSerGlnAlaArgHisValGlyLysLeuValLeuThrMetProProVal 1720

Qy    1359 TACAAT---------------ATCGAGCATGGCGGG------------------------ 1339
           ::::::               :::      ||||||
Db    1721 TrpAspAlaAlaGlyThrValLeuValThrGlyGlyThrGlyAlaLeuGlyAlaGluVal 1740

Qy    1338 ---CAGCACCTTACGGTCGAGCGAGGGGAGGTAGTCTTTGTACTTGTCGTCAAGCGATTT 1282
              :::||||||   :::|||||||||               ||||||||    :::|||
Db    1741 AlaArgHisLeuValIleGluArgGlyValArgAsnLeuValLeuValSerArgArgGly 1760

Qy    1281 GAGGATACCGGCATGAGCATCAGGATTTGTAGCCAATGCGTTGGC--------------- 1237
                          :::      :::   :::|||
Db    1761 ProAlaAlaSerGlyAlaAlaGluLeuValAlaGlnLeuThrAlaTyrGlyAlaGluVal 1780

Qy    1236 ------AAACTGTGCAAAACGAACCACCTCGGTACCACCGAAGAGCGTCTCGCTCAAATA 1183
                 :::   |||                              |||   ||||||::::::
Db    1781 SerLeuGlnAlaCysAspVal------------AlaAspArgGluThrLeuAlaLysVal 1796

Qy    1182 AGTCATCTCACGGTTGGCCTTGGCTCCTTCCTTATAAGCCTTTTCGAGAGAAGACAATAC 1123
                                         |||   |||:::   :::
Db    1797 ------------------------------LeuAlaSerIleProAspGluHisProLeu 1806

Qy    1122 ATCGCCATAGACAGCACTCTTGCCGTTCTTACGGATCCAGTCTGCGAATGCTCTTTCCTC 1063
              |||:::                  ||||||   |||   |||   |||
Db    1807 ThrAlaValValHisAlaAlaGlyValLeuAspAspGlyValSerGlu------------ 1822

Qy    1062 GGCACGCTTACGACCTATCACGTC---------------------------------AAG 1036
              :::||||||         :::                                   :::
Db    1823 ---SerLeuThrValGluArgLeuAspGlnValLeuArgProLysValAspGlyAlaArg 1841

Qy    1035 ACGAGCGAGACCGCGGTTCATACCGATCGAATTCTTCCAATAGTTAGCACTCTGAGCATA 976
                          :::      |||      |||   :::   |||::::::|||
Db    1842 AsnLeuLeuGluLeuIleAspProAspValAlaLeuValLeuPheSerSerValSerGly 1861

Qy     975 CTTGCTGGCATATTTGATACGGGTAGCCTGATCTGCGCTCATGGCTTCCTTCCAGATGCC 916
           :::   |||            |||:::         |||:::   ||||||   ||||||
Db    1862 ValLeuGlySerGlyGlyGlnGlyAsnTyrAlaAlaAlaAsnSerPheLeu---AspAla 1880

Qy     915 TTGCTT---------GATACCGCGAACTTCGATACGAGGATTGTTCTC------------ 877
           |||               :::            ||||||   :::
Db    1881 LeuAlaGlnGlnArgGlnSerArgGlyLeuProThrArgSerLeuAlaTrpGlyProTrp 1900

Qy     877 ------------------------------------------------------------ 877

Db    1901 AlaGluHisGlyMetAlaSerThrLeuArgGluAlaGluGlnAspArgLeuAlaArgSer 1920

Qy     876 ------------------------------------GTTTTCGATACGATCTTCCACACC 853
                                                  ||||||
Db    1921 GlyLeuLeuProIleSerThrGluGluGlyLeuSerGlnPheAspAlaAlaCysGlyGly 1940

Qy     852 CCAAGAAGTGAGGTAGCG---ATCCGTACTGCCCGGGAAACCGATGGTCATGGCATAGTC 796
                 :::   ||||||   :::|||   :::|||   :::||||||:::   |||
Db    1941 AlaHisThrValValAlaProValArgPheSerArgLeuSerAspGlyAsnAlaIleLys 1960

Qy     795 GTCAGCCTTGTAGCCTTGCATGGATACGGCAGCGAAGTAAACGGGCTTATAGGGTTTATT 736
              |||:::                              :::   ||||||:::|||
Db    1961 PheSerVal---------------------------LeuGlnGlyLeuValGly------ 1969

Qy     735 GTCCTTGCTGTATTCGGCCGGCCGGTTGTCGGCACCGGCATACACGCGGAATACGCTGAA 676

Db    1969 ------------------------------------------------------------ 1969

Qy     675 GTCGCCCGTGTGACGCGGCCACATCCAGTTGTCCGTATCGCCTCCGAACTTACCTACAGA 616
                             ||||||   |||   :::   |||:::
Db    1970 ------------------ProHisArgValAsnLysAlaAlaThr--------------- 1978

Qy     615 GCTGGGAGGAGCAAATACCATACGAACGTCCTTGAATACATCGTAGACGATGAGGAAGTA 556
                                                        ||||||   |||
Db    1979 ------------------------------------------AlaAspAspAlaGluSer 1984

Qy     555 TTCGTTGTTGGAATAGAAAGGCTCTACGATGCAGAGTTGGTTCTCGTCTGCATTTTCTTT 496
                       :::   ||||||   |||||||||               :::   :::
Db    1985 LeuArgLysArgLeuGlyArgLeuProAspAlaGluGlnHisArgIleLeuLeuAspLeu 2004

Qy     495 TTTGGCCAGTTCTTGGCATACCTCCTGAGCTTTGCGCAGACGCTCCATCTCGTCAGTGAT 436
                       :::|||   :::|||   ||||||   :::      :::   :::|||
Db    2005 ValArgMetHisValAlaAlaValLeuGlyPheAlaGlySerGlnGluIleThrAlaAsp 2024

Qy     435 ---ACCCTTGAGCTGTCCTTCTACCTTGTCCGTTACCTTCACGATCTTGCGCAGATACTT 379
              |||   ::::::         :::            :::   :::|||         :::
Db    2025 GlyThrPheLysValLeuGlyPheAspSerLeuThrValValGluLeuArgAsnArgIle 2044

Qy     378 CACGGAAAGACCCGGAATCGGAAGCTCCTCACCCATCGTGCGAGAAACGAAACCATCGCG 319
           :::|||   |||
Db    2045 AsnGlyAlaThr------------------------------------------------ 2048

Qy     318 CAGATAGTCGTGATCCACCGTGCTTTGGCTCTGGATAGCACCGTATCCGCAGTGGTGGTT 259
                                   |||   |||   :::|||:::
Db    2049 ---------------------GlyLeuArgLeuProAlaThrLeuValPheAsnTyrPro 2061

Qy     258 GGTAAAGATCAGGCCCTGATCGGACACTGTGATACCGGTACATCCGCCACCGAAGATAAC 199
                 |||   ||||||         |||      |||      ||||||
Db    2062 ThrProAsp---AlaLeuAlaAlaHisLeuValThrAlaLeuSerAla------------ 2076

Qy     198 CACGGCATTGGCAATGGACGGCTTGTCGAAACTGTAGAGCGAATCCAACGG--------- 148
                             ||||||   |||   |||   :::   |||
Db    2077 ---------------AspArgLeuAlaGlyThrPheGluGluLeuAspArgTrpAlaAla 2091

Qy     147 ---------------------------GAGCGTAAAGCCGAGCTCACGCATTCGATCCAG 115
                                            :::|||::::::|||      :::|||
Db    2092 AsnLeuProThrLeuAlaArgAspGluAlaThrArgAlaGlnIleThrThrArgLeuGln 2111

Qy     114 ATTCTCCTG 106
              :::|||
Db    2112 AlaIleLeu 2114
```

RESULT 12
US-09-370-700-3
; Sequence 3, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;  LENGTH: 2152
;  TYPE: PRT
;  ORGANISM: Saccharopolyspora spinosa
US-09-370-700-3

Alignment Scores:
Pred. No.:              0.0213        Length:         2152
Score:                  119.00        Matches:        161
Percent Similarity:     30.96%        Conservative:   100
Best Local Similarity:  19.10%        Mismatches:     274
Query Match:            3.01%         Indels:         308
DB:                     4             Gaps:           33

US-10-008-355-1 (1-2139) x US-09-370-700-3 (1-2152)

Qy   2139 TTAGATCAACTTCAGCTCTTGGATGAGACGGGGGCACTGACCCCATTTGTCAATCATGAA 2080
          ||||||   |||   |||:::||||||      |||      |||      :::   |||
Db   1415 LeuAspGlyLeuAlaLeuValAspGluProThrAlaThrAlaProLeuGlyAspGlyGlu 1434

Qy   2079 GAGAACGTAGCGGATGTCCACGCT---------------GATTGTGCGCTGCAGATCGGG 2035
                :::|||         |||               |||      :::   :::|||
Db   1435 ValArgIleAlaMetArgAlaAlaGlyValAsnPheArgAspAlaLeuIleAlaLeuGly 1454

Qy   2034 TTCGAACTCGATGTCACCACTCATAGCTTCCCAGTTGCCATCGAAAGCAAGACCGATCAG 1975
                                     :::|||                  ::::::
Db   1455 Met------------------------TyrProGlyValAlaSerLeuGlySerGluGly 1466

Qy   1974 ACGGCCGTTCTTATCGAATAC------GGGGCTACCGGAGTTACCGCCCCGTGATGTCGTT 1921
                |||::::::|||         |||   ||||||:::         |||
Db   1467 AlaGlyValValValGluThrGlyProGlyValThrGlyLeuAlaProGlyAsp------ 1484

Qy   1920 GTTCGATAGGAAAGCGATATGGAGCTGACCGTTCTCGGCATAGCGACCATAGTTTTTGGT 1861
                                        |||:::|||:::            ||||||
Db   1485 ---------------------------ArgValMetGlyMetIleProLysAlaPheGly 1495

Qy   1860 GCGGAAGAGGTCGAGGATATTCTCCTGTACGGCAAACTC--------------------- 1822
                   |||   |||      :::      ::::::
Db   1496 ProLeuAlaValAlaAspHisArgMetValThrArgIleProAlaGlyTrpSerPheAla 1515

Qy   1821 ------------ATCGCTCTTAGGATCCTGCTTCTCCAATACGCCCTTGCCTGTCGTATG 1774
                      :::   :::      |||         |||||||||
Db   1516 ArgAlaAlaSerValProIleValPheLeuThrAlaTyrTyrAlaLeuValAspLeuAla 1535

Qy   1773 ATAGTTGTACCAGGCACCGTCCTGCGGTTCATATCCCTTGATGGAGCCGTAGCTCATACG 1714
             :::   ||||||      |||         |||      |||   |||
Db   1536 GlyLeuArgProGlyGluSerLeuLeuValHisSerAlaAlaGlyGlyValGly------ 1553

Qy   1713 CATGGTGAAGTTGGCATCGCTCGGCAGAGCACGTCCGGGGTACATCTCACGCAAACCGGC 1654
                   :::   :::||||||         :::   |||:::   |||   :::
Db   1554 MetAlaAlaIleGlnLeuAlaArgHisLeuGlyAlaGluValTyrAlaThrAlaSerGlu 1573

Qy   1654 ------------------------------------------------------------ 1654

Db   1574 AspLysTrpGlnAlaValGluLeuSerArgGluHisLeuAlaSerSerArgThrCysAsp 1593

Qy   1653 ---AAAGAAAAGACGCTTGCCCTTCTCAATGGCATAGGCATTGGCCATCGCATCGGCCTG 1597
             :::::::::   |||            |||   |||:::         :::   |||
Db   1594 PheGluGlnGlnPheLeuGlyAlaThrGlyGlyArgGlyValAsp------ValValLeu 1611

Qy   1596 AATAGCGCGAGCAGCAGCTATTACGCTCTTGGAAAGCTCTACTGCCGGATCTTTCTCGAT 1537
          ||||||   :::                                       |||   |||
Db   1612 AsnSerLeuAlaGlyGlu---------------------------------PheAlaAsp 1620

Qy   1536 AGC------------------CTTGGCAAACTTTTCCTTGTCCATGGA------------ 1507
          :::                     |||      ||||||      |||
Db   1621 AlaSerLeuArgMetLeuProArgGlyGlyArgPheLeuGluLeuGlyLysThrAspVal 1640

Qy   1507 ------------------------------------------------------------ 1507

Db   1641 ArgAspProValGluValAlaAspAlaHisProGlyValSerTyrGlnAlaPheAspThr 1660

Qy   1506 CTTGAGCATGGCATGGAACTTGTCGCTATAAGGAACCACACTCTTGTCGAA---TACGAA 1450
          :::|||   |||   :::   :::            |||   |||||||||   :::|||
Db   1661 ValGluAlaGlyProGlnArgIleGlyGluMetLeuHisGluLeuValGluLeuPheGlu 1680

Qy   1449 GTCTGCATACTT---------------------------CTTCGTGTCGCCTTTGAATTT 1417
             :::|||                              :::|||   |||   |||
Db   1681 GlyArgValLeuGluProLeuProValThrAlaTrpAspValArgGlnAlaProGluAla 1700

Qy   1416 CTTGTCGATTACATT---CTTGAATATATCGGGGAGCTTGTCGGCAGGGATACGCCGGCG 1360
          |||                     ::::::|||:::||||||            |||
Db   1701 LeuArgHisLeuSerGlnAlaArgHisValGlyLysLeuValLeuThrMetProProVal 1720

Qy   1359 TACAAT---------------ATCGAGCATGGCGGG------------------------ 1339
          ::::::               :::      ||||||
Db   1721 TrpAspAlaAlaGlyThrValLeuValThrGlyGlyThrGlyAlaLeuGlyAlaGluVal 1740

Qy   1338 ---CAGCACCTTACGGTCGAGCGAGGGGAGGTAGTCTTTGTACTTGTCGTCAAGCGATTT 1282
             :::||||||   :::|||||||||            |||||||||   :::|||
Db   1741 AlaArgHisLeuValIleGluArgGlyValArgAsnLeuValLeuValSerArgArgGly 1760

Qy   1281 GAGGATACCGGCATGAGCATCAGGATTTGTAGCCAATGCGTTGGC--------------- 1237
                         :::      :::   :::|||
Db   1761 ProAlaAlaSerGlyAlaAlaGluLeuValAlaGlnLeuThrAlaTyrGlyAlaGluVal 1780

Qy   1236 ------AAACTGTGCAAAACGAACCACCTCGGTACCACCGAAGAGCGTCTCGCTCAAATA 1183
                :::   |||                           |||   ||||||::::::
Db   1781 SerLeuGlnAlaCysAspVal------------AlaAspArgGluThrLeuAlaLysVal 1796

Qy   1182 AGTCATCTCACGGTTGGCCTTGGCTCCTTCCTTATAAGCCTTTTCGAGAGAAGACAATAC 1123
                                        |||   |||:::   :::
Db   1797 ------------------------------LeuAlaSerIleProAspGluHisProLeu 1806

Qy   1122 ATCGCCATAGACAGCACTCTTGCCGTTCTTACGGATCCAGTCTGCGAATGCTCTTTCCTC 1063
          |||:::                  ||||||   |||   |||   |||
Db   1807 ThrAlaValValHisAlaAlaGlyValLeuAspAspGlyValSerGlu------------ 1822

Qy   1062 GGCACGCTTACGACCTATCACGTC---------------------------------AAG 1036
             :::||||||         :::                                 :::
Db   1823 ---SerLeuThrValGluArgLeuAspGlnValLeuArgProLysValAspGlyAlaArg 1841

Qy   1035 ACGAGCGAGACCGCGGTTCATACCGATCGAATTCTTCCAATAGTTAGCACTCTGAGCATA 976
                         :::      |||      |||   :::   |||::::::|||
Db   1842 AsnLeuLeuGluLeuIleAspProAspValAlaLeuValLeuPheSerSerValSerGly 1861

Qy    975 CTTGCTGGCATATTTGATACGGGTAGCCTGATCTGCGCTCATGGCTTCCTTCCAGATGCC 916
          :::   |||            |||:::         |||:::   ||||||   ||||||
Db   1862 ValLeuGlySerGlyGlyGlnGlyAsnTyrAlaAlaAlaAsnSerPheLeu---AspAla 1880

Qy    915 TTGCTT---------GATACCGCGAACTTCGATACGAGGATTGTTCTC------------ 877
          |||               :::            ||||||   :::
Db   1881 LeuAlaGlnGlnArgGlnSerArgGlyLeuProThrArgSerLeuAlaTrpGlyProTrp 1900

Qy    877 ------------------------------------------------------------ 877

Db   1901 AlaGluHisGlyMetAlaSerThrLeuArgGluAlaGluGlnAspArgLeuAlaArgSer 1920

Qy    876 ---------------------------------------GTTTTCGATACGATCTTCCACACC 853
                                                 ||||||
Db   1921 GlyLeuLeuProIleSerThrGluGluGlyLeuSerGlnPheAspAlaAlaCysGlyGly 1940

Qy    852 CCAAGAAGTGAGGTAGCG---ATCCGTACTGCCCGGGAAACCGATGGTCATGGCATAGTC 796
                :::   ||||||   :::|||   :::|||   :::||||||:::   |||
Db   1941 AlaHisThrValValAlaProValArgPheSerArgLeuSerAspGlyAsnAlaIleLys 1960

Qy    795 GTCAGCCTTGTAGCCTTGCATGGATACGGCAGCGAAGTAAACGGCTTATAGGGTTTATT 736
             |||:::                           :::   ||||||:::|||
Db   1961 PheSerVal---------------------------LeuGlnGlyLeuValGly------ 1969

Qy    735 GTCCTTGCTGTATTCGGCCGGCCGGTTGTCGGCACCGGCATACACGCGGAATACGCTGAA 676

Db   1969 ------------------------------------------------------------ 1969

Qy    675 GTCGCCCGTGTGACGCGGCCACATCCAGTTGTCCGTATCGCCTCCGAACTTACCTACAGA 616
```

```
                         ||||||  |||   :::   |||:::
Db   1970 -----------------ProHisArgValAsnLysAlaAlaThr-------------- 1978

Qy    615 GCTGGGAGGAGCAAATACCATACGAACGTCCTTGAATACATCGTAGACGATGAGGAAGTA 556
                                                 ||||||   |||
Db   1979 -----------------------------------------AlaAspAspAlaGluSer 1984

Qy    555 TTCGTTGTTGGAATAGAAAGGCTCTACGATGCAGAGTTGGTTCTCGTCTGCATTTTCTTT 496
                      :::   ||||||   |||||||||           :::   :::
Db   1985 LeuArgLysArgLeuGlyArgLeuProAspAlaGluGlnHisArgIleLeuLeuAspLeu 2004

Qy    495 TTTGGCCAGTTCTTGGCATACCTCCTGAGCTTTGCGCAGACGCTCCATCTCGTCAGTGAT 436
                         :::|||   :::|||   ||||||   :::      :::   :::|||
Db   2005 ValArgMetHisValAlaAlaValLeuGlyPheAlaGlySerGlnGluIleThrAlaAsp 2024

Qy    435 ---ACCCTTGAGCTGTCCTTCTACCTTGTCCGTTACCTTCACGATCTTGCGCAGATACTT 379
             |||   ::::::      :::            :::   :::|||            :::
Db   2025 GlyThrPheLysValLeuGlyPheAspSerLeuThrValValGluLeuArgAsnArgIle 2044

Qy    378 CACGGAAAGACCCGGAATCGGAAGCTCCTCACCCATCGTGCGAGAAACGAAACCATCGCG 319
          :::|||   |||
Db   2045 AsnGlyAlaThr------------------------------------------------ 2048

Qy    318 CAGATAGTCGTGATCCACCGTGCTTTGGCTCTGGATAGCACCGTATCCGCAGTGGTGGTT 259
                                  |||   |||   :::|||:::
Db   2049 --------------------GlyLeuArgLeuProAlaThrLeuValPheAsnTyrPro 2061

Qy    258 GGTAAAGATCAGGCCCTGATCGGACACTGTGATACCGGTACATCCGCCACCGAAGATAAC 199
                |||   ||||||      |||      |||      ||||||
Db   2062 ThrProAsp---AlaLeuAlaAlaHisLeuValThrAlaLeuSerAla----------- 2076

Qy    198 CACGGCATTGGCAATGGACGGCTTGTCGAAACTGTAGAGCGAATCCAACGG--------- 148
                           ||||||      |||   |||   :::   |||
Db   2077 ---------------AspArgLeuAlaGlyThrPheGluGluLeuAspArgTrpAlaAla 2091

Qy    147 --------------------------GAGCGTAAAGCCGAGCTCACGCATTCGATCCAG 115
                                    :::|||::::::|||      :::|||
Db   2092 AsnLeuProThrLeuAlaArgAspGluAlaThrArgAlaGlnIleThrThrArgLeuGln 2111

Qy    114 ATTCTCCTG 106
             :::|||
Db   2112 AlaIleLeu 2114
```

RESULT 13
US-07-728-215-33
; Sequence 33, Application US/07728215
; Patent No. 5962643
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Dean
; APPLICANT: Quaranta, Vito
; APPLICANT: Pytela, Robert
; TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/728,215
; FILING DATE: 19910711
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P31 8717
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 846 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-728-215-33

```
Alignment Scores:
Pred. No.:               0.0227       Length:        846
Score:                   115.50       Matches:       119
Percent Similarity:      30.99%       Conservative:  66
Best Local Similarity:   19.93%       Mismatches:    201
Query Match:             3.02%        Indels:        211
DB:                      2            Gaps:          26
```

US-10-008-355-1 (1-2139) x US-07-728-215-33 (1-846)

```
Qy    642 GGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGG 701
          |||   ||||||         |||      |||               |||   :::
Db    282 GlyAsnLeuAsp------AlaProGluGlyGlyPheAspAlaIleMetGlnAlaIleAla 299

Qy    702 TGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTT 761
          ||||||         |||      :::|||   :::                  ||||||||
Db    300 CysArgSerGlnIleGlyTrpArgGluGlnAlaArg--------------ArgLeuLeu 314

Qy    762 CGCTGCCGTATCCATGCAAGGC------TACAAGGCTGACGACTATGCCATGACCATCGG 815
                            ||| |||      |||      |||                :::
Db    315 ValPheSerThrAspAla-GlyPheHisTyrAlaGlyAspGlyLysLeuGlyGlyValIl 334

Qy    816 TTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAA 875
                |||         :::   :::|||:::      |||
Db    334 eAlaProAsnAspGlyGluCysHisLeuSerProLysGlyGluTyrThrHisSerThrLe 354

Qy    876 CGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCAT 935
            ::::::   |||   |||                  |||   |||   :::      |||::
Db    354 uGlnAspTyrProSerIleSer------GlnIleAsnGlnLysValLysAspAsnAlaIl 372

Qy    936 GAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTA 995
            ::::         |||                  |||   |||   :::            ::
Db    372 eAsnIleIlePheAlaValThrAlaSerGlnLeuSerValTyrGluLysLeuValGluHi 392

Qy    996 TTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAA 1055
            :   :::   |||                      |||:::||||||
Db    392 sIleGlnGlySer--------------SerAlaAlaLysLeuAsp------------- 402

Qy   1056 GCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTC-- 1113
                                                 |||      |||   |||
Db    403 ---------------------------------------AsnAspSerSerAsnValVa 409

Qy   1114 ------------------TATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGA 1154
                            |||      :::   ||||||:::|||         |||::
Db    409 lGluLeuValLysGluGluTyrArgLysIleSerSerSerValGlu------MetLysAs 427

Qy   1155 AGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGA 1214
            :   |||         :::   ::::::||||||   |||         |||         |||   ||
Db    427 pAsnAlaThrGlyAspValLysIleThrTyrPheSerSerCysLeuSerAsnGlyProGl 447

Qy   1215 GGTG-----------------------------------GTTCGTTTT---GCACA 1232
          ||||                                   |||   |||   |||||
Db    447 uValGlnThrSerLysCysAspAsnLeuLysGluGlyGlnGlnValSerPheThrAlaGl 467

Qy   1233 GTTTG-----------------------CCAACGCATTGGCTACAAATCCTGATGCTCA 1268
          | :::                       |||   |||   |||   :::   :::
Db    467 n-IleGlnLeuLeuLysCysProGluAspProArgAspTrpThrGlnThrIleHisIleS 487

Qy   1269 TGCCGGTATCCTCAAATCGCTTG------------------------------------ 1291
            ||||||         |||   :::
Db    487 erProValGlyIleAsnGluValMetGlnIleGlnLeuThrMetLeuCysSerCysProC 507
```

```
Qy   1291 ------------------------------------------------------------ 1291

Db    507 ysGluAsnProGlySerIleGlyTyrGlnValGlnAlaAsnSerCysSerGlyHisGlyT 527

Qy   1292 --------------------------------------------------------ACGA 1295
                                                                  :
Db    527 hrSerMetCysGlyIleCysAsnCysAspAspSerTyrPheGlyAsnLysCysGluCysS 547

Qy   1296 CAAGTACAAAGACTACCTCC--------------------------------CCTCGCT 1322
           ::::|||    ||||||                                  |
Db    547 erAlaThrAspLeuThrSerLysPheAlaAsnAspThrSerCysArgAlaAspSerThrS 567

Qy   1323 CGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAA 1382
           |||||     ||| :::                ||| |||     |||    :::
Db    567 erThrThrAspCys-SerGlyArgGlyHisCysCysValGlyAlaCysGluCysHisLys 586

Qy   1383 GCTCCCCGATATATTCAA--------------------GAATGTAATCGACAAGAAATT 1421
             |||     |||:::                      ||||||:::
Db    587 ArgProAsnProIleGluIleIleSerGlyLysHisCysGluCysAspAsnPheSerCys 606

Qy   1422 CAAAGGCGACACGAAGAAGTATGC--------------------------------AGA 1448
           :::|||       :::::|||                                 |||
Db    607 GluArgAsnArgAsnGlnLeuCysSerGlyProAspHisGlyThrCysGluCysGlyArg 626

Qy   1449 CTTCGTATTCGA--------------CAAGAGTGTGGTTCCTTATAGCGACAAGTTCCA 1493
              :::   :::                      ||||||
Db    627 CysLysCysLysProGlyTrpThrGlySerAsnCysGlyCysGlnGluSerAsnAspThr 646

Qy   1494 TGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGC 1553
           |||               |||   |||:::::|||   |||:::
Db    647 Cys------MetProProGlyGlyGlyGluIleCysSerGlyHis--------------- 659

Qy   1554 AGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGAT--- 1610
                                                       |||   |||:::
Db    660 ---------------------------------------------GlyThrCysGluCys 664

Qy   1611 GGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGAT 1670
           |||   |||   |||          |||       |||       |||
Db    665 GlyValCysLysCysThrValAsnAspGlnGlyArgPheSerGlyArgHisCysGluLys 684

Qy   1671 GTACCCCGGACGTGCTCTGCCGAG-----------------------CGATGCCAACTT 1706
              |||   |||||||||                             :::||||||:::
Db    685 CysPro---ThrCysSerGlyArgCysGlnGluLeuLysAspCysValGlnCysGlnMet 703

Qy   1707 CACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGC------ 1760
           :::         ||||||:::          |||          |||      |||
Db    704 TyrLysThrGlyGluLeuLys------AsnGlyAspAspCysAlaArgAsnCysThrGln 721

Qy   1761 CTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGA 1820
              |||   :::    ::::::       :::   |||   :::|||    :::
Db    722 PheValProValGlyValGluLys-ValGluIleAspGluThrLysAspGluGlnMetCy 741

Qy   1821 TGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTA 1880
            :::|||      :::::::::            :::|||                  :::||
Db    741 sLysPhePheAspGluAspAspCysLysPheMetPhe------------------LysTy 755

Qy   1881 TGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGAC-------------- 1926
           |:::|||   |||:::|||||||:::          |||    :::
Db    755 rSerGluGlnGlyGluLeuHisValTyrAlaGlnGluAsnLysGluCysProAlaLysVa 775

Qy   1927 -------------ATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCG 1973
                           |||      |||          :::   |||
Db    775 lPheMetLeuGlyIleValMetGlyValIleAlaAlaIleVal------------------ 789

Qy   1974 TCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGT 2016
            |||:::|||||||||          |||:::   ::::::
Db    790 -LeuValGlyLeuAlaIleLeuLeuLeuTrpLysLeuLeuThr 803
```

RESULT 14
US-08-938-085A-33
; Sequence 33, Application US/08938085A
; Patent No. 6339148
; GENERAL INFORMATION:
;   APPLICANT: Sheppard, Dean
;   APPLICANT: Quaranta, Vito
;   APPLICANT: Pytela, Robert
;   TITLE OF INVENTION: A No. 6339148el Integrin Beta Subunit and Uses
;   TITLE OF INVENTION: Thereof
;   NUMBER OF SEQUENCES: 62
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Townsend and Townsend and Crew LLP
;   STREET: Two Embarcadero Center, Eighth Floor
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94111-3834
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/938,085A
;   FILING DATE: 26-SEP-1997
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/728,215
;   FILING DATE: 11-JUL-1991
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Parent, Annette S.
;   REGISTRATION NUMBER: 42,058
;   REFERENCE/DOCKET NUMBER: 02307O-080210US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 576-0200
;   TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 33:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 846 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-938-085A-33

```
Alignment Scores:
Pred. No.:              0.0227        Length:          846
Score:                  115.50        Matches:         119
Percent Similarity:     30.99%        Conservative:    56
Best Local Similarity:  19.93%        Mismatches:      201
Query Match:            3.02%         Indels:          211
DB:                     4             Gaps:            26

US-10-008-355-1 (1-2139) x US-08-938-085A-33 (1-846)

Qy    642 GGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGG 701
           |||   ||||||          |||       |||             |||    :::
Db    282 GlyAsnLeuAsp------AlaProGluGlyGlyPheAspAlaIleMetGlnAlaIleAla 299

Qy    702 TGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTT 761
           ||||||         |||       :::|||    :::                |||||||||
Db    300 CysArgSerGlnIleGlyTrpArgGluGlnAlaArg---------------ArgLeuLeu 314

Qy    762 CGCTGCCGTATCCATGCAAGGC------TACAAGGCTGACGACTATGCCATGACCATCGG 815
                           |||·|||      |||      |||                :::
Db    315 ValPheSerThrAspAla-GlyPheHisTyrAlaGlyAspGlyLysLeuGlyGlyValIl 334

Qy    816 TTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAA 875
              |||          :::   :::|||:::        |||
Db    334 eAlaProAsnAspGlyGluCysHisLeuSerProLysGlyGluTyrThrHisSerThrLe 354

Qy    876 CGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCAT 935
           ::::::    |||    |||          |||    |||    :::          |||::
Db    354 uGlnAspTyrProSerIleSer------GlnIleAsnGlnLysValLysAspAsnAlaIl 372

Qy    936 GAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTA 995
```

```
              ::::         |||                        |||    |||    :::         ::
Db      372 eAsnIleIlePheAlaValThrAlaSerGlnLeuSerValTyrGluLysLeuValGluHi 392

Qy      996 TTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAA 1055
            :   ::::   |||                        |||:::||||||
Db      392 sIleGlnGlySer---------------SerAlaAlaLysLeuAsp-------------- 402

Qy     1056 GCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTC-- 1113
                                                   |||      |||   |||
Db      403 ----------------------------------------AsnAspSerSerAsnValVa 409

Qy     1114 ------------------TATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGA 1154
                              |||     :::  ||||||:::|||        |||::
Db      409 lGluLeuValLysGluGluTyrArgLysIleSerSerSerValGlu------MetLysAs 427

Qy     1155 AGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGA 1214
            :   |||        :::   ::::::||||||  |||      |||       |||  ||
Db      427 pAsnAlaThrGlyAspValLysIleThrTyrPheSerSerCysLeuSerAsnGlyProGl 447

Qy     1215 GGTG--------------------------------------GTTCGTTTT---GCACA 1232
            ||||                                      |||   |||   |||||
Db      447 uValGlnThrSerLysCysAspAsnLeuLysGluGlyGlnGlnValSerPheThrAlaGl 467

Qy     1233 GTTTG-----------------------CCAACGCATTGGCTACAAATCCTGATGCTCA 1268
            | :::                       |||   |||  |||   :::   :::
Db      467 n-IleGlnLeuLeuLysCysProGluAspProArgAspTrpThrGlnThrIleHisIleS 487

Qy     1269 TGCCGGTATCCTCAAATCGCTTG------------------------------------ 1291
              ||||||     |||    :::
Db      487 erProValGlyIleAsnGluValMetGlnIleGlnLeuThrMetLeuCysSerCysProC 507

Qy     1291 ------------------------------------------------------------ 1291

Db      507 ysGluAsnProGlySerIleGlyTyrGlnValGlnAlaAsnSerCysSerGlyHisGlyT 527

Qy     1292 ---------------------------------------------------------ACGA 1295
                                                                       :
Db      527 hrSerMetCysGlyIleCysAsnCysAspAspSerTyrPheGlyAsnLysCysGluCysS 547

Qy     1296 CAAGTACAAAGACTACCTCC-------------------------------CCTCGCT 1322
            :::::|||     ||||||                                       |
Db      547 erAlaThrAspLeuThrSerLysPheAlaAsnAspThrSerCysArgAlaAspSerThrS 567

Qy     1323 CGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAA 1382
            |||||     ||| :::                |||       |||       |||    :::
Db      567 erThrThrAspCys-SerGlyArgGlyHisCysCysValGlyAlaCysGluCysHisLys 586

Qy     1383 GCTCCCCGATATATTCAA--------------------GAATGTAATCGACAAGAAATT 1421
            .|||      |||:::                      ||||||:::
Db      587 ArgProAsnProIleGluIleIleSerGlyLysHisCysGluCysAspAsnPheSerCys 606

Qy     1422 CAAAGGCGACACGAAGAAGTATGC--------------------------------AGA 1448
            :::|||            :::::::|||                               |||
Db      607 GluArgAsnArgAsnGlnLeuCysSerGlyProAspHisGlyThrCysGluCysGlyArg 626

Qy     1449 CTTCGTATTCGA---------------CAAGAGTGTGGTTCCTTATAGCGACAAGTTCCA 1493
               :::   :::                      ||||||
Db      627 CysLysCysLysProGlyTrpThrGlySerAsnCysGlyCysGlnGluSerAsnAspThr 646

Qy     1494 TGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGC 1553
            |||                |||   |||:::::::|||   |||:::
Db      647 Cys------MetProProGlyGlyGlyGluIleCysSerGlyHis--------------- 659

Qy     1554 AGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGAT--- 1610
                                                        |||    |||:::
Db      660 ---------------------------------------------GlyThrCysGluCys 664

Qy     1611 GGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGAT 1670
            |||   |||   |||                |||      |||      |||
Db      665 GlyValCysLysCysThrValAsnAspGlnGlyArgPheSerGlyArgHisCysGluLys 684

Qy     1671 GTACCCCGGACGTGCTCTGCCGAG-----------------------CGATGCCAACTT 1706
               |||   |||||||||                               :::||||||:::

Db      685 CysPro---ThrCysSerGlyArgCysGlnGluLeuLysAspCysValGlnCysGlnMet 703

Qy     1707 CACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGC------ 1760
            :::              ||||||:::          |||         |||       |||
Db      704 TyrLysThrGlyGluLeuLys------AsnGlyAspAspCysAlaArgAsnCysThrGln 721

Qy     1761 CTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGA 1820
               |||   :::      ::::::      :::   |||   :::|||    :::
Db      722 PheValProValGlyValGluLys-ValGluIleAspGluThrLysAspGluGlnMetCy 741

Qy     1821 TGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTA 1880
               :::|||      ::::::::            :::|||                 :::||
Db      741 sLysPhePheAspGluAspAspCysLysPheMetPhe-----------------LysTy 755

Qy     1881 TGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGAC------------- 1926
            |:::|||   |||:::||||||:::                  |||     :::
Db      755 rSerGluGlnGlyGluLeuHisValTyrAlaGlnGluAsnLysGluCysProAlaLysVa 775

Qy     1927 ------------ATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCG 1973
                          |||      |||         :::   |||
Db      775 lPheMetLeuGlyIleValMetGlyValIleAlaAlaIleVal----------------- 789

Qy     1974 TCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGT 2016
               |||:::|||||||||               |||:::   ::::::
Db      790 -LeuValGlyLeuAlaIleLeuLeuLeuTrpLysLeuLeuThr 803
```

```
RESULT 15
US-08-484-105-14
; Sequence 14, Application US/08484105
; Patent No. 5589341
;  GENERAL INFORMATION:
;    APPLICANT:  STILLMAN, Bruce
;    APPLICANT:  BELL, Stephen P
;    APPLICANT:  KOBAYASHI, Ryuji
;    APPLICANT:  RINE, Jasper
;    APPLICANT:  FOSS, Margit
;    APPLICANT:  McNALLY, Francis J
;    APPLICANT:  LAURENSON, Patricia
;    APPLICANT:  HERSKOWITZ, Ira
;    APPLICANT:  LI, Joachim J
;    APPLICANT:  GAVIN, Kimberly
;    TITLE OF INVENTION:  ORIGIN OF REPLICATION COMPLEX GENES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,105
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59032/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  885 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
US-08-484-105-14

Alignment Scores:
Pred. No.:              0.0264        Length:         885
Score:                  115.00        Matches:        158
Percent Similarity:     30.61%        Conservative:   108
Best Local Similarity:  18.18%        Mismatches:     316
Query Match:            3.01%         Indels:         287
DB:                     1             Gaps:           34

US-10-008-355-1 (1-2139) x US-08-484-105-14 (1-885)

Qy     90 CCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCTCCC 149
          |||:::|||   :::   |||:::::         :::            |||
Db     28 ProArgArgArgArgArgGlyAspAlaThrGluHisGluValIleAsnLeuValArgTyr 47

Qy    150 GTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTTCGG 209
             |||      |||             :::   |||:::          |||
Db     48 AspGlyValArgLeuTyrProGlyValThrIleValCysLysValGluGlyAlaAspGlu 67

Qy    210 TGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCACTG 269
                   |||      |||   ||||||                 |||          :::
Db     68 LeuSerAlaTyrMetIleHisGluValArgLeuAsnThrSerAsnTyr---------Val 84

Qy    270 CGGATACGGTGC------------------------------------------------ 281
             :::   |||
Db     85 GluLeuTrpCysLeuAsnTyrLeuSerTrpTyrGluIleAsnAlaAlaGluArgTyrLys 104

Qy    282 -----------------TATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGA 323
                            |||         |||      |||   ::: :::         ::
Db    105 GlnLeuAspGlyGluPheTyrGluThrAsnLysGluLysGlyAspLys-PhePheGluGl 124

Qy    324 TGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGT---CTTTCCGTGAA 380
          :   |||   |||::::::::        ||||||   :::         ||||||
Db    124 uThrPheAlaSerGlnSerIleLysAsnGluLeuTyrLeuThrAlaGluLeuSerGluIl 144

Qy    381 GTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCAC 440
           ||||||||   :::      |||   :::                        |||
Db    144 eTyrLeuArgAspLeuGlnPheValAlaAsn-----------------------IleLy 156

Qy    441 TGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAG-----------------------GT 476
           :::|||   |||   |||          |||                         ::
Db    156 sAsnGluLysGluTyrLeuAspSerValAsnGluGlyLysMetAspSerAsnMetPheLe 176

Qy    477 ATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATC-------- 528
          :|||:::                          :::      |||   |||
Db    176 uCysArgSerAlaCysLeuProSerGlyThrAsnLeuAlaAspLeuAspIleHisPhePh 196

Qy    529 -GTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGA 587
             |||      |||:::|||               :::   |||         :::||
Db    196 eGluGluLysIleArgSerSerAsnProLysValSerLeuGluTyr------LeuArgAs 214

Qy    588 CGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAA 647
          |:::   :::
Db    214 pIleThrLeu-------------------------------------------------- 217

Qy    648 CTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGA 707
                      |||:::             :::      :::|||:::|||
Db    218 ---------ProLysLeuProLysProLeuAsnLysSerLysValHisAlaArgGluLy 234

Qy    708 CAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGC 767
                    :::             ||||||   |||   |||       :::
Db    234 sValValAlaThrLysLeuGlnSerAspAsnThrProSerLysLysSerPheGlnGlnTh 254

Qy    768 CGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATC-------------- 813
           ||||||            :::            |||   ||||||
Db    254 rValSerLysThrAsnAlaGluValGlnArgIleAlaSerThrIleValAsnGluLysGl 274

Qy    814 -----------------------------GGTTTCCCGGGCAGTACGGATCGCTACCT 842
                                       :::      |||   :::   :::
Db    274 uAlaIleSerAspAsnGluSerAspLeuSerGluTyrHisGluSerLysGluGluPheAl 294

Qy    843 CACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAA----- 897
             :::         :::         |||   |||:::               |||
Db    294 aAsnAlaSerSerSerAspSerAspGluGluPheGluAspTyrGlnSerAlaGluGluLe 314

Qy    898 ---------------------------GTTCGCGGTATCAAGCAAGGCATCTGGAAGGA 929
                                     ||||||   ||||||         |||
Db    314 uAlaIleValGluProAlaLysLysLysValArgSerIleLysProAspIle------Pr 332

Qy    930 AGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGC 989
                           |||   ::::::   ::::::   ::::::
Db    332 oIleSerProValLysSerGlnThrProLeuGlnProSerAlaValHisSerSerProAr 352

Qy    990 TAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGG 1049
           ::::::||||||:::|||
Db    352 gLysPhePheLysAsnAsnIle------------------------------------- 359

Qy   1050 TCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGT------AAGAACGGCAA 1103
               ||||||   ::::::|||:::        :::   :::      ||||||   ||
Db    360 ----ValArgAla---LysLysAlaTyrThrProPheSerLysArgTyrLysAsnProLy 377

Qy   1104 GAGTGCTGTCTATGGCGATGTA-------------------------TTGTCTTCTCT 1136
          |          |||:::                             :::::::||
Db    377 sIleProAspLeuAsnAspIlePheGlnArgHisAsnAsnAspLeuAspIleAlaAlaLe 397

Qy   1137 CGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGAC 1196
          ||||:::   ::::::         |||      :::|||            ||||
Db    397 uGluGluArgPheArgThr---ValSerAlaLysGlyLysMet-----------GluTh 412

Qy   1197 GCTCTTCGGTGGTACC--------------------------------GAGGTGGTTCG 1223
          |:::|||                                         |||:::|||::
Db    412 rIlePheSerLysValLysLysGlnLeuAsnSerArgAsnSerLysGluGluIleValLy 432

Qy   1224 TTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAA 1283
          :   |||   |||   |||   |||                 :::      |||   |||
Db    432 sAlaAlaAspPheAspAsnTyrLeuProAlaArgGluAsnGluPheAlaSerIleTyrLe 452

Qy   1284 ATCGCTT------------------------------------------------- 1290
           ||||||
Db    452 uSerLeuTyrSerAlaIleGluAlaGlyThrSerThrSerIleTyrIleAlaGlyThrPr 472

Qy   1291 ------------------------GACGACAAGTACAAAGACTACCTCCCCTCGCTCGA 1325
                                  :::      ||||||   :::   |||   ||
Db    472 oGlyValGlyLysThrLeuThrValArgGluValValLysAspLeuMetThrSerAlaAs 492

Qy   1326 CCGTAAGGTGCTGCCCGCCATG--------------------CTCGATATTGTACGCCG 1364
          |:::|||   ||||||                          |||   ||||||:::
Db    492 pGlnLysGluLeuProArgPheGlnTyrIleGluIleAsnGlyLeuLysIleValLys-- 511

Qy   1365 GCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATC------GACAAGAA 1418
                     |||         ::::::|||         |||         :::|||
Db    512 ---------AlaSerAspSerTyrGluValPheTrpGlnLysIleSerGlyGluLysLe 528

Qy   1419 ATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTA 1478
              |||         :::      :::|||   |||:::|||         ||||||
Db    528 uThrSerGlyAlaAlaMetGluSerLeuGluPheTyrPheAsnLys------ValPro-- 545

Qy   1479 TAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTAT 1538
                                                                   |||
Db    546 --------------------------------------------------------AlaTh 547

Qy   1539 CGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCA 1598
           :::|||   |||   |||   |||         :::   |||                  ||
Db    547 rLysLysArgProIleValValLeuLeuAspGluLeuAspAlaLeuValSerLysSerGl 567

Qy   1599 GGCCGATGCGATGGCCAATGCCTAT---------GCCATTGAGAAGGGCAAGCGTCTTTT 1649
          |   |||   |||   |||   :::                              |||   :::
Db    567 n---AspValMetTyrAsnPhePheAsnTrpAlaThrTyrSerAsnAlaLysLeuIleVa 586

Qy   1650 CTTTGCC--------------GGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAG 1694
             |||                  |||   |||   :::   |||         :::   ||
Db    586 lValAlaValAlaAsnThrLeuAspLeuProGluArgHisLeuGlyAsnLysIleSerSe 606

Qy   1695 CGATGCCAACTTCACC-------------------------------------------- 1710
```

```
                 |          ||||||
Db       606 rArgIleGlyPheThrArgIleMetPheThrGlyTyrThrHisGluGluLeuArgThrIl 626

Qy      1711 -------ATGCGTATGAGCTACGGCTCCATCAAGGGATAT------GAACCGCAGGACGG 1757
                    :::|||:::   |||   :::         :::     :::|||:::   ||
Db       626 eIleAsnLeuArgLeuLysTyrLeuAsnGluSerSerPheTyrValAspProGluThrGl 646

Qy      1758 TGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAG 1817
             |:::   |||                     :::         ::::::   |||
Db       646 ySerSerTyrMetIleSerProAspSerSerThrIleGluThrAspGluGluGluLysAr 666

Qy      1818 CGATGAGTTTGCCGTACAGGAGAATATCCTC----------------------------- 1848
                :::|||:::      :::   :::
Db       666 gLysAspPheSerAsnTyrLysArgLeuLysLeuArgIleAsnProAspAlaIleGluIl 686

Qy      1849 ------------------------------GACCTCTTCCGC-----------ACCAA 1865
                                           |||:::   |||            ||
Db       686 eAlaSerArgLysIleAlaSerValSerGlyAspValArgArgAlaLeuLysValValLy 706

Qy      1866 AAACTATGGTCGCTATGCCGAGAACGGTCAGCTC---------------CATATCGCTTT 1910
             |            ||||||||||||        |||
Db       706 sArgAlaValGluTyrAlaGluAsnAspTyrLeuLysArgLeuArgTyrGluArgLeuVa 726

Qy      1911 CCTATCGAACAACGACATCACGGGCGGTAACTCCGGT---------------------- 1947
                 |||      |||   :::|||      :::|||
Db       726 lAsnSerLysLysAspThrSerGlyAsnGlyThrGlyAsnGluGluLeuGlnSerValGl 746

Qy      1948 ------------------------------AGCCCCGTATTCGATAAGAACGGCCGTCT 1976
                                           |||            :::         ::
Db       746 uIleLysHisIleThrLysAlaLeuAsnGluSerSerThrSerProGluGlnGlnPheIl 766

Qy      1977 GATCGGTCTTGCTTTCGATGGC-------------------------------------- 1998
             :    ||||||:::|||   |||
Db       766 eSerGlyLeuSerPheSerGlyLysLeuPheLeuTyrAlaLeuIleAsnLeuIleLysLy 786

Qy      1999 ----AACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAAT 2054
                       :::         |||||||||      |||       |||         ::
Db       786 sLysGlnThrAspValGlnLeuGlyAspIleValGluGluMetArgLeuLeuIleAspVa 806

Qy      2055 CAGCGTGGACATCCGCTACGTTCTC 2079
             ::::   :::   :::|||:::|||
Db       806 lAsnGlyAsnAsnLysTyrIleLeu 814
```

Search completed: May 16, 2003, 12:39:14
Job time : 73.5 secs

D P

GenCore version 5.1.4_p5_4578

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 16, 2003, 12:35:41 ; Search time 83.5 Seconds
(without alignments)
4941.515 Million cell updates/sec

Title: US-10-008-355-1
Perfect score: 3820
Sequence: 1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 362588 seqs, 96450795 residues

Total number of hits satisfying chosen parameters: 725176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
```
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110341_9460/app_query.fasta_1.2311
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10008355_@CGN_1_1_25_@runat_16052003_110341_9460
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3719	97.4	712	9	US-10-008-355-2	Sequence 2, Appli
	2	3652	95.6	699	9	US-10-008-355-8	Sequence 8, Appli
	3	1276	33.4	720	9	US-10-008-355-9	Sequence 9, Appli
	4	907.5	23.8	716	9	US-10-008-355-7	Sequence 7, Appli
	5	884.5	23.2	734	9	US-10-008-355-5	Sequence 5, Appli
	6	807.5	21.1	732	9	US-10-008-355-6	Sequence 6, Appli
	7	274	7.2	52	9	US-10-008-355-3	Sequence 3, Appli
	8	148.5	3.9	1528	9	US-09-945-917-3	Sequence 3, Appli
	9	148.5	3.9	1583	9	US-09-945-917-4	Sequence 4, Appli
	10	145	3.8	1367	10	US-09-801-368-108	Sequence 108, App
	11	133	3.5	883	10	US-09-815-242-13382	Sequence 13382, A
	12	127	3.3	883	10	US-09-815-242-13684	Sequence 13684, A
	13	122.5	3.2	26	9	US-10-008-355-18	Sequence 18, Appl
	14	118.5	3.1	2042	9	US-10-192-584-6	Sequence 6, Appli
	15	118	3.1	1814	10	US-09-920-552-103	Sequence 103, App
	16	117	3.1	708	10	US-09-925-301-1012	Sequence 1012, Ap
	17	116.5	3.0	1169	10	US-09-801-368-106	Sequence 106, App
	18	116	3.0	1138	12	US-10-036-328A-6	Sequence 6, Appli
	19	116	3.0	1210	12	US-10-036-328A-2	Sequence 2, Appli
	20	115.5	3.0	846	9	US-10-072-841-33	Sequence 33, Appl
c	21	115.5	2.9	1270	9	US-10-101-464A-979	Sequence 979, App
	22	114	3.0	2701	9	US-10-171-311-83	Sequence 83, Appl
	23	113	3.0	687	9	US-09-969-384-27	Sequence 27, Appl
	24	112	2.9	1014	9	US-10-097-340-4	Sequence 4, Appli
	25	111.5	2.9	682	10	US-09-920-552-13	Sequence 13, Appl
	26	111.5	2.9	1023	9	US-09-884-696-5	Sequence 5, Appli
	27	111	2.9	1114	10	US-09-782-874-2	Sequence 2, Appli
	28	111	2.9	5877	9	US-10-142-515-11	Sequence 11, Appl
	29	110	2.9	2039	9	US-10-192-584-7	Sequence 7, Appli
	30	109.5	2.9	845	9	US-10-027-806-62	Sequence 62, Appl
	31	109.5	2.9	845	9	US-10-034-623-62	Sequence 62, Appl
	32	109.5	2.9	845	9	US-10-027-801-62	Sequence 62, Appl
	33	109	2.9	630	9	US-10-213-990-45	Sequence 45, Appl
	34	109	2.9	2382	9	US-10-196-935A-2	Sequence 2, Appli
c	35	107.5	2.7	890	9	US-10-101-464A-958	Sequence 958, App
	36	107.5	2.8	1394	9	US-09-945-917-57	Sequence 57, Appl
	37	107.5	2.8	1449	9	US-09-945-917-58	Sequence 58, Appl
	38	107	2.8	889	9	US-09-909-567B-44	Sequence 44, Appl
	39	107	2.8	1607	10	US-09-938-275-10	Sequence 10, Appl
	40	106.5	2.8	523	10	US-09-815-242-11918	Sequence 11918, A
	41	106	2.8	556	10	US-09-801-368-426	Sequence 426, App
	42	106	2.8	972	9	US-10-100-049-23	Sequence 23, Appl
	43	105.5	2.8	5179	9	US-10-025-380-1068	Sequence 1068, Ap
	44	105.5	2.8	5179	10	US-09-922-217-1068	Sequence 1068, Ap
	45	105.5	2.8	5179	10	US-09-833-263-1068	Sequence 1068, Ap

1A

ALIGNMENTS

RESULT 1
US-10-008-355-2
; Sequence 2, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 712
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-2

Alignment Scores:

Pred. No.:	7.63e-310	Length:	712
Score:	3719.00	Matches:	712
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.36%	Indels:	0

```
DB:                  9          Gaps:        0

US-10-008-355-1 (1-2139) x US-10-008-355-2 (1-712)

Qy      1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20

Qy     61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAsp 40

Qy    121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60

Qy    181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 SerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80

Qy    241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100

Qy    301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuPro 120

Qy    361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140

Qy    421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160

Qy    481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180

Qy    541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200

Qy    601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro 220

Qy    661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240

Qy    721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln 260

Qy    781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280

Qy    841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300

Qy    901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320

Qy    961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340

Qy   1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360

Qy   1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380

Qy   1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400

Qy   1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420

Qy   1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSer 440

Qy   1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAsp 460

Qy   1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480

Qy   1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500

Qy   1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520

Qy   1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540

Qy   1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560

Qy   1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580

Qy   1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600

Qy   1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620

Qy   1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640

Qy   1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660

Qy   1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680

Qy   2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGly 700

Qy   2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATC 2136
          ||||||||||||||||||||||||||||||||||||
```

```
Db     701 GlnCysProArgLeuIleGlnGluLeuLysLeuIle 712

RESULT 2
US-10-008-355-8
; Sequence 8, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 699
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-8

Alignment Scores:
Pred. No.:              4.19e-304    Length:         699
Score:                  3652.00      Matches:        699
Percent Similarity:     100.00%      Conservative:   0
Best Local Similarity:  100.00%      Mismatches:     0
Query Match:            95.60%       Indels:         0
DB:                     9            Gaps:           0

US-10-008-355-1 (1-2139) x US-10-008-355-8 (1-699)

Qy       1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20

Qy      61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAsp 40

Qy     121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60

Qy     181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 SerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80

Qy     241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100

Qy     301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuPro 120

Qy     361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140

Qy     421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160

Qy     481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180

Qy     541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200

Qy     601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro 220

Qy     661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240

Qy     721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln 260

Qy     781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280

Qy     841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300

Qy     901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320

Qy     961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340

Qy    1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360

Qy    1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380

Qy    1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400

Qy    1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420

Qy    1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSer 440

Qy    1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAsp 460

Qy    1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480

Qy    1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500

Qy    1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520

Qy    1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540

Qy    1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560
```

```
Qy   1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580

Qy   1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600

Qy   1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620

Qy   1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640

Qy   1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660

Qy   1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680

Qy   2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGG 2097
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrp 699

RESULT 3
US-10-008-355-9
; Sequence 9, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-9

Alignment Scores:
Pred. No.:              2.07e-100      Length:          720
Score:                  1276.00        Matches:         283
Percent Similarity:     56.53%         Conservative:    128
Best Local Similarity:  38.93%         Mismatches:      286
Query Match:            33.40%         Indels:          30
DB:                     9              Gaps:            11

US-10-008-355-1 (1-2139) x US-10-008-355-9 (1-720)

Qy     16 AAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGGGTAGCCAAAGCCGAC 75
          |||   :::||||||             :::|||             :::|||  ||||||
Db      3 LysArgLeuLeuLeuProLeuPheAlaValLeuCysLeuCysGlnIleAlaHisAlaAsp 22

Qy     76 AAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTC 135
          :::||||||||||||:::   :::|||   ::::::      :::|||:::|||
Db     23 GluGlyMetTrpLeuMetGlnGlnLeuGlyArgLys---TyrAlaGlnMetLysGluArg 41

Qy    136 GGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCC 195
          |||      :::      ||||||:::   :::         |||:::   :::|||
Db     42 GlyLeuLysMetLysGluTyrAspLeuTyrAsnProAsnGlyThrSerLeuLysAspAla 61

Qy    196 GTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTT 255
          ||||||:::|||  ||||||||||||||        |||||||||:::||||||:::
Db     62 ValValLeuPheAspGlyGlyCysThrGlyGluValValSerAspArgGlyLeuValLeu 81

Qy    256 ACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTAT 315
          ||||||||||||||||||||      |||||||:::   |||||||::::::|||:::|||
Db     82 ThrAsnHisHisCysGlyTyrAspMetIleGlnAlaHisSerThrLeuGluHisAsnTyr 101

Qy    316 CTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCC 375
          |||   :::||||||   :::           :::|||||||||      :::|||
Db    102 LeuGluAsnGlyPheTrpAlaMetArgGluAlaAspGluLeuProAsnLysAspIleSer 121

Qy    376 GTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGT 435
          |||   ::::::   ||||||      |||||||||   |||:::      ||||||
Db    122 ValValPheIleAspLysIleGluAspValThrAspTyrValLysLysAspLeuLysAla 141

Qy    436 ATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAA 495
          |||   |||                     :::   :::      |||:::||||||
Db    142 IleLysAspProAsnSerMetAspTyrLeuSerProLysTyrLeuGlnLysLeuAlaAsp 161

Qy    496 AAAGAA---------------AATGCAGACGAGAACCAACTCTGCATC---GTAGAGCCT 537
          |||:::               :::|||         |||    :::   ::::::
Db    162 LysLysAlaGlyLysAsnPheSerAlaLysAsnProGlyLeuSerValGluIleLysAla 181

Qy    538 TTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATG 597
          ||||||      |||   |||      :::            :::   |||||||||:::
Db    182 PheTyrGlyGlyAsnLeuTyrLeuMetPheThrLysLysThrTyrThrAspValArgLeu 201

Qy    598 GTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGG 657
          |||   |||||||||:::|||:::||||||||||||   |||||||||||||||:::|||
Db    202 ValGlyAlaProProThrSerIleGlyLysPheGlyAlaAspThrAspAsnTrpIleTrp 221

Qy    658 CCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCG 717
          ||||||||||||||||||||||||:::||||||:::||||||         :::      |||
Db    222 ProArgHisThrGlyAspPheSerIlePheArgIleTyrAlaAspLysAsnGlyAsnPro 241

Qy    718 GCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATG 777
          |||   ||||||:::||||||   |||      ||||||      |||      :::|||:::
Db    242 AlaProTyrSerGluAspAsnValProLeuLysProLysArgPhePheAsnIleSerLeu 261

Qy    778 CAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGC 837
             |||   :::   :::||||||||||||      :::||||||||||||:::|||   |||
Db    262 GlyGlyValGlnGluAsnAspTyrAlaMetIleMetGlyPheProGlyThrThrHisArg 281

Qy    838 TACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAA 897
          |||   |||:::      |||::::::            :::|||:::   ||||||
Db    282 TyrPheThrAlaSerGluValAspGluTrpLysSerIleAspAsnAspIleArgIleArg 301

Qy    898 GTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGT 957
          :::|||   |||:::||||||:::               |||   ||||||         :::
Db    302 MetArgAspIleArgGlnGlyValMetLeuArgGluMetLeuAlaAspProGlnIleLys 321

Qy    958 ATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATG 1017
          |||   |||::::::|||||||||   |||   |||   :::|||   :::||||||
Db    322 IleMetTyrSerAlaLysTyrAlaAlaSerGlnAsnAlaTyrLysArgAlaIleGlyAla 341

Qy   1018 AACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTC 1077
          |||      :::         :::      |||:::|||   :::
Db    342 AsnTrpAlaIleLysThrArgGlyLeuArgGlnAsnLysGlnAlaMetGlnAspArgLeu 361

Qy   1078 GCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTC 1137
                |||   |||   |||   :::         |||   :::   :::   :::
Db    362 IleAlaTrpGlyAlaLysGlnGly---ThrProArgTyrGluGluAlaValHisGluIle 380

Qy   1138 GAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACG 1197
          :::               :::   |||      |||            :::   |||
Db    381 AspAlaThrValAlaLysArgAlaAspLeuArgArgArgTyrTrpMetIleGluGluGly 400

Qy   1198 CTCTTCGGTGGTACCGAGGTGGTTCGT------------TTTGCACAGTTTGCCAACGCA 1245
          :::         |||   |||         |||                        |||
Db    401 IleIleArgGlyIleGluPheAlaArgSerProIleProThrGluAspGluThrLysAla 420

Qy   1246 TTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTAC--- 1302
          |||         |||   ||||||   |||               |||         ||||||
```

```
Db    421 LeuGlnGlyAsn---AspAlaSerAla------ArgLysGluAlaIleAspLysIleArg 437

Qy   1303 -----------------------AAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTG 1338
                                 |||||||||          :::|||:::||||||
Db    438 ThrArgTyrSerLysPheAlaAsnLysAspTyrSerAlaGluValAspLysLysValAla 457

Qy   1339 CCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTC 1398
             |||||||||                :::   ||||||   :::   ||||||   :::
Db    458 ValAlaMetLeuThrGluTyrLeuLysGluIleProTyrGluAsnLeuPro---LeuHis 476

Qy   1399 AAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTC 1458
                :::::: :::|||  ||||||  :::   |||  |||  :::|||
Db    477 LeuArgLeuValLysAspArgPheAlaGlyAspValGlnAlaTyrValAspAspIlePhe 496

Qy   1459 GACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAG 1518
             :::||||||               :::|||  |||
Db    497 AlaArgSerValPheGlySerGluAlaGlnPheAspAlaPheAlaAlaValProSerVal 516

Qy   1519 GAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATA 1578
          ||||||   |||               :::||||||  |||    :::  ||||||
Db    517 GluLysLeuAla-----------GluAspProMetValLeuPheAlaSerSerValPhe 532

Qy   1579 GCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGC 1638
                   |||   :::      :::                     |||   :::
Db    533 AspGluTyrArgLysLeuTyrAsnGluLeuArgProTyrAspAspProIleLeuArgAla 552

Qy   1639 AAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGAT 1698
          :::|||   :::   |||||||||   ||||||          :::            |||
Db    553 GlnArgThrTyrIleAlaGlyLeuLeuGluMetAspGlyAspGlnAspGlnPheProAsp 572

Qy   1699 GCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGT 1758
          ||||||   |||:::|||   :::||||||   :::|||||||||   |||:::|||
Db    573 AlaAsnLeuThrLeuArgPheThrTyrGlyGlnValLysGlyTyrSerProArgAspAsn 592

Qy   1759 GCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGC 1818
             :::|||   :::   ||||||      ||||||:::||||||:::||||||   :::
Db    593 ValTyrTyrGlyHisGlnThrThrLeuAspGlyValMetGluLysGluAspProAspAsn 612

Qy   1819 GATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGC 1878
             ||||||   |||         :::      ::::::      |||::::::||||||
Db    613 TrpGluPheValValAspProLysLeuLysAlaValTyrGluArgLysAspPheGlyArg 632

Qy   1879 TATGCCGAG---AACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGC 1935
          ||||||:::   :::|||::::::   :::                        ||||||
Db    633 TyrAlaAspArgSerGlyArgMetProValAlaPheCysAlaThrThrHisThrThrGly 652

Qy   1936 GGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGAT 1995
          ||||||||||||||||||||     :::   ||||||   ||||||||||||    ||||||
Db    653 GlyAsnSerGlySerProValMetAsnAlaAsnGlyGluLeuIleGlyLeuAsnPheAsp 672

Qy   1996 GGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATC 2055
             |||||||||   :::   ||||||||||::::::         |||   ||||||:::|||
Db    673 ArgAsnTrpGluGlyValGlyGlyAspIleGlnTyrLeuAlaAspTyrGlnArgSerIle 692

Qy   2056 AGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTC 2115
             ||||||||||||||||||||   :::|||||||||   |||   |||   ||||||
Db    693 IleValAspIleArgTyrValLeuLeuValIleAspLysValGlyGlyCysGlnArgLeu 712

Qy   2116 ATCCAAGAGCTGAAGTTGATC 2136
          :::   |||:::   ::::::
Db    713 LeuAspGluMetAsnIleVal 719
```

RESULT 4
US-10-008-355-7
; Sequence 7, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-008-355-7

```
Alignment Scores:
Pred. No.:               8.02e-69      Length:          716
Score:                   907.50        Matches:         239
Percent Similarity:      49.21%        Conservative:    135
Best Local Similarity:   31.45%        Mismatches:      291
Query Match:             23.76%        Indels:          95
DB:                      9             Gaps:            21

US-10-008-355-1 (1-2139) x US-10-008-355-7 (1-716)

Qy      1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
          |||:::   |||  |||:::   :::   |||   |||:::           |||
Db      1 MetArgPheAsnLeuLeuSerLeuSerValLeuAlaThrLeuIleThrValAspSer--- 19

Qy     61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
             |||   :::|||||||||:::      :::|||      |||
Db     20 ---ThrHisAlaGlyGluGlyMetTrpValProGlnGlnLeu---ProGluIleAlaGly 37

Qy    121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
             :::::::::   |||      |||      :::   |||   :::
Db     38 ProLeuLysGlnAlaGlyLeuGlnLeuSerProGluGlnLeuSerAsnLeuThrGlyAsp 57

Qy    181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
             :::            |||   :::      |||   ||||||         ||||||
Db     58 ProMetGlyAlaValValSerLeu------GlyAsnCysThrAlaSerLeuValSerPro 75

Qy    241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
          :::||||||:::   ||||||||||||||   ||||||||||||||      ||||||
Db     76 GluGlyLeuValIleThrAsnHisHisCysAlaTyrGlyAlaIleGlnLeuAsnSerThr 95

Qy    301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
             :::   :::::::::||||||   :::   |||      :::|||:::
Db     96 ProLysLysAsnLeuIleLysGluGlyPheAsnAlaLeuThrGlnAlaAspGluValSer 115

Qy    361 ATT---CCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTA 417
          |||             :::      |||   :::|||      ||||||   :::
Db    116 AlaGlyProAsnAlaArgIleTyrValLeuGluGlnIleThrAspValThrAlaGlnAla 135

Qy    418 GAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTA 477
          :::      |||         ::::::      :::         |||   |||
Db    136 LysAlaAlaLeuAlaAlaAlaGlyAsnAspProPheLysArgThrThrAlaLeuGluThr 155

Qy    478 TGC------CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTA 531
                   ||||||:::||||||   |||      ::::::         |||
Db    156 PheSerLysGlnGluIleAlaLysCysGluGlu---GluGlnGlyTyrArgCysGlnPhe 174

Qy    532 GAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTT 591
                |||         |||   |||   :::               |||||||||
Db    175 PheSerPheAlaGlyGlyAsnThrTyrArgValPheLysAsnLeuGluIleLysAspVal 194

Qy    592 CGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGG 651
          |||:::|||:::||||||      ||||||||||||||||||||||||   |||||||||
Db    195 ArgLeuValTyrAlaProGlnGlySerValGlyLysPheGlyGlyAspValAspAsnTrp 214

Qy    652 ATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAAC 711
          ||||||||||||||||||||||||||||||||   :::|||   |||   |||   |||
Db    215 MetTrpProArgHisThrGlyAspPheSerPheTyrArgAlaTyrValGlyLysAspGly 234

Qy    712 CGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTA 771
          :::||||||   :::||||||:::|||   ||||||:::|||   ::::::
Db    235 LysProAlaSerPheSerLysGluAsnIleProTyrArgProLysHisTrpLeuLysPhe 254
```

```
Qy     772 TCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACG 831
Db     255 SerAspGlnProLeuGlyAspGlyAspPheValMetValAlaGlyTyrProGlyArgThr 274

Qy     832 GATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGT 891
Db     275 AsnArgTyrAlaLeuVal------------------AlaGluPheGluAsnThrAlaHis 288

Qy     892 ATCGAAGTTCGCGGTATCAAGCAA------------------GGCATCTGGAAGGAAGCC 933
Db     289 TrpThrTyrProValIleGlyGlnHisPheLysAsnLeuIleAlaLeuIleGluAlaAla 308

Qy     934 ATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAAC 993
Db     309 SerLysGlnAsnProAspIleGlnValLysTyrAlaSerThrLeuAlaGlyLeuAsnAsn 328

Qy     994 TATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGT 1053
Db     329 ThrSerLysAsnPheAspGlyGlnLeuAspGlyPheArgArgIleAsnAlaIleGlyGln 348

Qy    1054 AAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGC------------ 1101
Db     349 LysGlnSerGluGluThrAlaValLeuAlaTrpLeuLysGlnGlnGlyIleArgGlyHis 368

Qy    1102 AAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCC 1161
Db     369 GluAlaLeuAlaAlaHisGlnThrLeuValAspLeuThrGluGlnTyr------------ 384

Qy    1162 AAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTT 1221
Db     385 LysAlaAsnGlnAspArgAspPheValLeuGlyGln---PheAsnGlySerGlyValIle 403

Qy    1222 CGTTTTGCACAGTTTGCCAACGCATTGGCT---------ACAAATCCTGATGCTCAT--- 1269
Db     404 GlyValAlaValAsnLeuTyrArgLeuAlaIleGluArgThrLysSerAspAlaGlnArg 423

Qy    1270 ------------------------------------GCCGGTATCCTCAAATCGCTTGAC 1293
Db     424 GluAlaGlyTyrGlnGluArgAspLeuProThrIleGluGlyAsnLeuLysGlnMetGlu 443

Qy    1294 GACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTG------------------ 1335
Db     444 ---------ArgArgTyrLeuProGluMetAspArgGlnMetGlnGlnTyrTrpLeuThr 460

Qy    1336 ------------CTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCC------ 1377
Db     461 GluTyrAsnLysLeuPro---------------ValLysGlnArgValAlaAlaIleAsp 475

Qy    1378 ------------GACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAA 1425
Db     476 ValTrpLeuGlyAspGlyIleProAlaThrLeuLysArgLeuGlyAspThrLysLeuSer 495

Qy    1426 GGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGAC 1485
Db     496 SerSerGluGluArg--------------------------------------------- 500

Qy    1486 AAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAA 1545
Db     501 ---------------LeuLysTrpPheAsnAlaAspArgAlaAlaPheGluSerSerGln 515

Qy    1546 GATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGAT 1605
Db     516 AspProAlaIleArgTyrAlaValAlaIleMetProAlaLeuLeuGluIleGluArgGln 535

Qy    1606 GCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGT 1665
Db     536 AsnLysIleArgThrGlyGluLeuLeuLysAlaArgProIleTyrLeuGlnAlaLeuAla 555

Qy    1666 GAGATGTACCCC------GGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATG 1719
Db     556 AspTyrAsnLysSerHisGlyLysPheValTyrProAspAlaAsnSerSerLeuArgIle 575

Qy    1720 AGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACG 1779
Db     576 ThrPheGlyHisValLysGlyTyrSerProLysAspGlyValGluTyrThrProPheThr 595

Qy    1780 ACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAG 1839
Db     596 ThrLeuGlnGlyValMetAlaLys---AsnThrGlyValGluProPheAspSerProLys 614

Qy    1840 AATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAAC------GGT 1893
Db     615 SerLeuIleAsnAlaIleLysAlaLysSerTyrAlaAsnLeuAlaAspGlnArgIleGly 634

Qy    1894 CAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCC 1953
Db     635 ThrValProValAsnPheLeuSerAspLeuAspIleThrGlyGlyAsnSerGlySerPro 654

Qy    1954 GTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATG 2013
Db     655 ValLeuAspAlaHisGlyLysLeuValGlyLeuAlaPheAspGlyAsnTrpGluSerVal 674

Qy    2014 AGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTAC 2073
Db     675 SerSerAsnTrpValPheAspProValMetThrArgThrIleAlaValAspSerArgTyr 694

Qy    2074 GTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTG 2133
Db     695 ValGlnTrpIleMetThrGluValAlaProAlaProHisLeuLeuLysGluLeuAsnLeu 714
```

RESULT 5
US-10-008-355-5
; Sequence 5, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
US-10-008-355-5

Alignment Scores:

Pred. No.:	7.62e-67	Length:	734
Score:	884.50	Matches:	247
Percent Similarity:	48.58%	Conservative:	129
Best Local Similarity:	31.91%	Mismatches:	280
Query Match:	23.15%	Indels:	118
DB:	9	Gaps:	25

US-10-008-355-1 (1-2139) x US-10-008-355-5 (1-734)

```
Qy       6 AATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGGGTAGC 65
Db      15 AsnThrMetLysLysTrp-LeuLeuSerValAla------ValAlaAlaSerPheAlaSe 32

Qy      66 CAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAAT 125
Db      32 rHisAlaAspGluGlyMetTrpGlnProHisGlnLeu---ProAlaMetAlaAspValLe 51

Qy     126 GCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCAT 185
Db      51 uLysAlaLysGlyLeuGluIleAspAlaLysSerIleSerLysLeuThrGluPheProMe 71

Qy     186 TGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGG 245
```

```
Db     71 t---AsnAlaValIleSerLeu---GlyGlyCysThrAlaSerPheValSerProLysGl 89

Qy    246 CCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGA 305
          ||||:::   ||||||||||||||    ||||||:::||||||       ||||||    ::
Db     89 yLeuValValThrAsnHisHisCysAlaTyrGlySerIleGlnTyrAsnSerThrProGl 109

Qy    306 TCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG---AT 362
          :  :::   |||:::|||||||||:::::::::|||        :::|||||||||
Db    109 uLysAsnLeuLeuGlnAspGlyPheLeuAlaLysThrPheAlaAspGluLeuProAlaAl 129

Qy    363 TCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGG 422
           ||||||      |||              :::      ||||||::::::|||:::
Db    129 aProGlySerArgValTyrValThrGluAspValThrAsnValThrGluArgValLysAl 149

Qy    423 ACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCA 482
              |||:::      |||                           :::|||     ||
Db    149 aGlyLeuGluAsnLysThr-----------------------GlyArgGluPheTyrGl 161

Qy    483 AGAACTGGCCAAAAAAGAAAATGCA--------------------GACGAGAACCAACT 521
          |   :::      :::|||   |||                       ||||||
Db    161 nGlyValGluAsnGlnGluLysAlaLeuValAlaGluCysGluLysAspGluGlyTyrAr 181

Qy    522 CTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATT 581
           |||   |||      |||:::           ||||||:::|||:::
Db    181 gCysGlnValTyrSerPheHisGlyGlyLeuGluTyrTyrLeuValLysGlnLeuGluIl 201

Qy    582 CAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATAC 641
           :::||||||||::::||||:::   |||      ||||||||||||:::||||||||
Db    201 eArgAspValArgLeuValTyrAsnProAlaGlySerValGlyLysTyrGlyGlyAspVa 221

Qy    642 GGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGG 701
           ||||||||||||||||||||||||||||||||:::|||     :::|||     |||
Db    221 lAspAsnTrpMetTrpProArgHisThrGlyAspTyrSerPheTyrArgAlaTyrValSe 241

Qy    702 TGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTT 761
                :::   :::||||||||||:::|||   ||||||   ||||||:::|||      ||
Db    241 rLysAsnGlyLysProAlaGluPheSerAlaAspAsnValProTyrGluProLysSerPh 261

Qy    762 CGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCC 821
          |      ||||||    :::|||             |||:::   |||      |||:::||
Db    261 eLeuLysValSerAlaLysGlyValSerGluGlyAspPheValMetValAlaGlyTyrPr 281

Qy    822 GGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAA------AA 875
          ||||   |||:::||||||        |||:::     |||::::::       |||
Db    281 oGlyArgThrAsnArgTyrArgThrAlaThrGluValGlnAsnGluPheGluTrpAlaTy 301

Qy    876 CGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCAT 935
                :::      |||   |||:::      |||   |||
Db    301 rProGluGlyLysMetLeuArgGluArgPheIleGlu------IleIleLysAlaThrAl 319

Qy    936 GAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTA 995
                            |||||||||||   |||:::   |||      |||||||
Db    319 aProGluGlySerAspGluArgIleLysTyrGluSerGlnIleAlaGlyLeuAlaAsnTy 339

Qy    996 TTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAA 1055
          |   ||||||            |||                :::        ::::::     ::
Db    339 rAlaLysAsnPheThrSerMetIleGluPheTyrGlyLysSerThrMetLeuAlaAspAr 359

Qy   1056 GCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAG---AACGGCAAGAGTGCTGT 1112
          ::::|||   |||   |||:::||||||   |||   :::   :::   |||
Db    359 gLysAlaLeuGluAlaLysLeuAlaGluTrpIleAlaLysAspSerSerArgGluAlaLy 379

Qy   1113 CTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCG 1172
           ||||||      |||:::   |||:::   |||              :::||||||::::
Db    379 sTyrGlyLysThrLeuAlaGluLeuAsp---AlaLeuIleAlaLysSerLysAlaHisGl 398

Qy   1173 TGAG------------ATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGT 1220
          :|||              ::::::|||:::|||   |||         |||
Db    398 nGluArgAspMetIleLeuSerTyrIleSerSerThrThrMetLeuProThr-------- 415

Qy   1221 TCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCT 1280
                           ||||||   |||               ||||||
Db    416 ----------------AlaAsnAsnLeuTyrArgLeu------AlaHisGlu-------- 425

Qy   1281 CAAATCGCTTGACGACAAGTACAAAGAC-------------------------------- 1308
           |||   |||   |||        ::::::
Db    426 -LysGlnLeuProAspMetGlnArgGluProGlyPheGlnAspArgAspMetThrArgPh 445

Qy   1309 ------------------------------TACCTCCCCTCGCTCGACCGTAAGGTGCT 1337
                                        |||      |||:::|||:::   |||||
Db    445 eLysAlaSerMetGluArgIleAspArgArgTyrAlaAlaSerValAspLysAlaValLe 465

Qy   1338 GCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATT 1397
           |     ||||||                           |||   :::||||||
Db    465 uPheAspMetLeuLysArgTyrAlaAlaLeuProGluAlaGlnArgLeuPro-------- 482

Qy   1398 CAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAG-------------------- 1437
                :::||||||   |||      |||   |||
Db    483 -------AlaMetAspLysAlaPheGlyIleAspAsnLysValAsnGluAlaLysLeuAl 500

Qy   1438 -AAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGAC---AAGTTCCA 1493
            |||      |||   ::::::   |||:::   :::         |||   :::
Db    500 aLysThrLeuAspLysMetTyrAlaLysThrGluLeuGlyAsnLysAspValArgLeuAl 520

Qy   1494 TGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGC 1553
                |||   ||||||:::|||         |||               ||||||||
Db    520 aTrpMetGluLysSerValAspAspPheLys------------AlaSerLysAspProPh 536

Qy   1554 AGTAGAGCTTTCCAAGAGCGTA-------------------------------------- 1575
           ::::::   :::   ::::::
Db    536 eIleGlnPheAlaValAlaMetTyrAspThrAsnMetSerGluGluLysLysGluLysGl 556

Qy   1576 ------------ATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTA 1622
                        :::         |||         ||||||:::         |||||
Db    556 uLeuAspGlyGluLeuMetLysValArgProGlnTyrMetAspAlaIleIle---AlaTy 575

Qy   1623 TGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACG 1682
           |   :::|||:::||||||   ::::::
Db    575 rAsnLeuGluGlnGlyLysProValTyr-------------------------------- 584

Qy   1683 TGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATA 1742
                    :::|||||||||   ::::::|||::::::   |||   :::||||||||
Db    585 ----------AlaAspAlaAsnSerSerLeuArgValThrValGlyHisValLysGlyTy 601

Qy   1743 TGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAA 1802
           |   |||:::||||||                  |||      :::|||:::::::::||
Db    601 rSerProLysAspGlyLeuValAlaValProPheThrArgLeuGluGlyIleValGlnLy 621

Qy   1803 GCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCAC 1862
           |   |||         |||   |||         :::      |||:::|||   :::
Db    621 s---AspThrGlyIleAspProPheAspAlaProLysGlnGlnLeuGluLeuIleLysGl 640

Qy   1863 CAAAAACTATGGTCGCTATGCCGAGAACGGT------CAGCTCCATATCGCTTTCCTATC 1916
           |||   ||||||   :::                  :::   :::   |||||||
Db    640 nLysGlnTyrGlyAspPheTyrMetLysSerIleAspSerValProValAsnPheLeuSe 660

Qy   1917 GAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCT 1976
           |      |||   |||||||||||||||||||||||         :::         ||
Db    660 rThrLeuAspThrThrGlyGlyAsnSerGlySerProThrLeuAsnGlyArgAlaGluLe 680

Qy   1977 GATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACC 2036
           |:::||||||   |||||||||   :::|||::::::   |||         |||:::
Db    680 uValGlyLeuLeuPheAspGlyValTyrGluSerIleIleGlyGlyTrpAlaPheAspAs 700

Qy   2037 CGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATG 2096
           ::::::   |||:::|||   ||||||   ||||||:::|||:::::::::
Db    700 nGluIleAsnArgSerIleHisValAspSerArgTyrMetLeuTrpValMetLysTyrLe 720

Qy   2097 GGGTCAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATC 2136
                         |||:::   |||::::::::::::
Db    720 uAspHisAlaAspAsnLeuLeuAlaGluMetGluIleVal 733
```

RESULT 6
US-10-008-355-6
; Sequence 6, Application US/10008355
; Patent No. US20020164759A1

```
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
;   LENGTH: 732
;   TYPE: PRT
;   ORGANISM: Shewanella putrefaciens
US-10-008-355-6

Alignment Scores:
Pred. No.:              3.04e-60      Length:         732
Score:                  807.50        Matches:        235
Percent Similarity:     44.81%        Conservative:   119
Best Local Similarity:  29.75%        Mismatches:     289
Query Match:            21.14%        Indels:         147
DB:                     9             Gaps:           25

US-10-008-355-1 (1-2139) x US-10-008-355-6 (1-732)

Qy     22 ATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGGGTAGCCAAAGCCGACAAAGGC 81
          |||   |||   |||||||||:::|||             |||:::|||  ||||||:::|||
Db      3 IleAlaLeuValAlaAlaLeuValLeu---ThrCysGlyIleAlaThrAlaAspGluGly 21

Qy     82 ATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTT 141
             |||         ::::::             |||::::::           |||
Db     22 GlnTrpGlnProTyrGlnMet---ProSerIleAlaAspLysLeuSerAlaArgGlyIle 40

Qy    142 ACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTT 201
             :::|||   |||   |||          |||      |||                 |||
Db     41 AspIleProAlaAspLysLeu------AlaAspLeuThrSerTyrProMetAsnAlaVal 58

Qy    202 ATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAAC 261
          :::      |||   ||||||            ||||||   |||||||||:::    ||||||
Db     59 ValGlyLeuGlyTyrCysThrAlaSerPheValSerProGlnGlyLeuValValThrAsn 78

Qy    262 CACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGC 321
          ||||||||   |||   |||||||||          :::         :::|||:::||||||
Db     79 HisHisCysAlaTyrLysAlaIleGlnTyrAsnThrLysLysGluHisAsnTyrLeuGlu 98

Qy    322 GATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAG 381
             ||||||::::::    :::|||   :::|||   |||          |||   :::    :::
Db     99 GlnGlyPheLeuAlaThrSerMetAspLysGlu---ProSerAlaGlyProAsnGluArg 117

Qy    382 ---TATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATC 438
             |||:::    :::          |||||||||               |||      :::
Db    118 LeuTyrIleThrGlu------AlaValThrAspValThrSerAspValThrLysAspLeu 135

Qy    439 ACT---GACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAA 495
          :::   |||   ::::::|||       :::    :::           :::   |||   |||
Db    136 SerGlnAspProLeuLysArgTyrGluGluIleGluAsnHisSerLysAlaLeuIleLys 155

Qy    496 AAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAA 555
                   |||||||:::|||       |||   |||      |||::::::        |||
Db    156 SerCysGluAlaAspAspAsnTyrArgCysAsnValArgSerPheHisAsnGlyLeuGlu 175

Qy    556 TACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGC 615
          |||:::||||||            :::     :::|||||||||:::|||:::|||||||||
Db    176 TyrTyrLeuIleLysGlnLeuMetIleArgAspValArgLeuValTyrAlaProProGlu 195

Qy    616 TCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGAC 675
          |||||||||    :::|||||||||    |||||||:::     :::|||||||||:::||||||
Db    196 SerValGlyGlyTyrGlyGlyAspIleAspAsnTyrGluTyrProArgHisSerGlyAsp 215

Qy    676 TTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGAC 735
```

```
          |||:::         |||   |||   |||   |||    :::||||||    :::|||:::|||
Db    216 PheAlaPheLeuArgAlaTyrValGlyLysAspGlyLysProAlaAlaPheSerGluAsp 235

Qy    736 AATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGAC 795
          |||   ||||||   |||         :::         |||:::        |||   ||||||
Db    236 AsnIleProTyrThrProLysSerTyrLeuLysValAsnAlaAspGlyValLysAlaGly 255

Qy    796 GACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC---CTCACTTCT--- 849
          |||            |||:::||||||:::|||:::||||||   |||||||||
Db    256 AspGlyValPheValAlaGlyTyrProGlyThrThrAsnArgTyrAsnLeuThrSerGlu 275

Qy    850 ------------------TGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGT 891
                            |||             :::   :::                  :::
Db    276 LeuLysPheAlaSerAspTrpLeuTyrProThrGlnAlaLysArgTyr---------Gln 292

Qy    892 ATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATG---AGCGCAGATCAG 948
          :::::::::         |||                  |||||||||         |||
Db    293 LeuGlnIleAspThrIle-----------------GluAlaMetGlyGlnLysAspAla 306

Qy    949 GCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCG 1008
                |||||||||||         |||      ||||||        |||
Db    307 AspIleAlaIleLysTyrAlaGlyAsnMetAlaSerMetAlaAsnArgMetLysLysLeu 326

Qy   1009 ATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAA 1068
             |||:::    |||              |||::::::|||   |||:::         |||
Db    327 AsnGlyLeuLeuAlaGlyPheLysAlaThrAspIleValGlyIleLysGlnGlnArgGlu 346

Qy   1069 AGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAG------------------------ 1104
             |||      |||:::   ||||||
Db    347 AsnAspPheLeuAlaTrpLeuThrLysAsnProAsnLeuAsnGlnAsnLeuIleSerGlu 366

Qy   1105 ---------------------------------------------------------AGT 1107
                                                                  :::
Db    367 LeuGluValLeuLeuAlaGluGlnGlnLeuGlnThrGlnThrAsnTyrTyrPheThrAsn 386

Qy   1108 GCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTAT---------------AAG 1152
          |||            :::|||:::::              |||               |||
Db    387 AlaGlnSerSerThrLeuLeuThrAlaAlaAsnAsnLeuTyrArgLeuAlaLysGluLys 406

Qy   1153 GAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACC 1212
          :::   :::   |||   ||||||:::   |||         |||
Db    407 GlnLysSerAspAlaGluArgGluIleGlyTyrGlnGluArgAspLeu------------ 422

Qy   1213 GAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCC 1272
                         |||   |||::::::
Db    423 ---------------AlaMetPheSerSerArg------------------------- 428

Qy   1273 GGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAG 1332
                ||||||   :::|||                        |||   |||   |||
Db    429 ------LeuLysArgIleAspSer--------------------SerPheAspValLys 439

Qy   1333 GTG-----------------------CTGCCCGCCATGCTCGATATTGTACGCCGGCGT 1368
          |||                       |||   |||   |||               |||
Db    440 ValAspLysThrLeuTrpLeuGlnAspLeuAsnAlaTyrLeuSerGlnProAsnArgVal 459

Qy   1369 ATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC 1428
                |||   :::               |||:::   ||||||
Db    460 AlaAlaLeuAspAsnMet------------LeuAsnLeuAsnAspLys------------ 471

Qy   1429 GACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAG 1488
                                        :::   |||:::                |||
Db    472 -----------------------------AsnValSerLeuAlaAlaLysLeuAspGly 481

Qy   1489 TTCCATGCCATGCTCAAGTCCATGGACAAG--------------------GAAAAGTTT 1527
             :::::::::         |||:::                          |||
Db    482 LeuTyrSerLeuThrThrLeuThrAspGlnAlaGlnArgLeuAlaTrpMetGluAlaAsp 501

Qy   1528 GCCAAGGCTATCGAG------AAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCT 1581
          |||||||||   |||         ||||||   :::   |||:::   ::::::
Db    502 AlaLysAlaPheGluThrSerSerAspProPheIleArgLeuAlaValAlaLeuTyrAsp 521

Qy   1582 GCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGCGAAG 1641
                |||   :::                        |||      :::         :::
```

```
Db     522 ThrAsnMetAlaGlnGluLysAlaGluLysIleLeuAlaGlyLysLeuSerThrAlaArg 541

Qy    1642 CGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCG------AGC 1695
                 :::   |||   :::   :::   |||               |||
Db     542 ProAlaTyrMetAlaAlaValIleAspTyrTyrLysAlaAsnAsnTrpProValTyrPro 561

Qy    1696 GATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGAC 1755
           |||||||||   |||:::|||:::|||||||||   :::   ||||||:::   :::|||
Db     562 AspAlaAsnGlyThrLeuArgIleSerTyrGlyMetValAspGlyTyrGlnSerArgAsp 581

Qy    1756 GGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAG 1815
                 :::            |||         |||::::::   |||
Db     582 AlaLeuTyrLysGlnProPheThrArgLeuAspGlyIleValAlaLysHis---ThrGly 600

Qy    1816 AGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGT 1875
              :::   :::         :::   :::||||||         :::   :::|||
Db     601 ValGluProTyrAsnAlaProLysLysLeuLeuAspAlaIleSerValGlnArgPheGly 620

Qy    1876 CGCTAT------------------------------------------------------ 1881
              :::
Db     621 AspHisLeuValLysSerValTyrGlnAspProArgGlyTrpIleCysArgLeuPheSer 640

Qy    1882 --------------GCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGAC 1926
                             |||         :::   :::   |||||||||:::   |||
Db     641 CysLeuAspLysProGluGluPheAsnSerValProValAsnPheLeuSerSerValAsp 660

Qy    1927 ATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTT 1986
              ||||||||||||||||||||||||||||||:::      |||   |||:::||||||
Db     661 ThrThrGlyGlyAsnSerGlySerProValPheAsnGlyLysGlyGluLeuValGlyLeu 680

Qy    1987 GCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAG 2046
              ||||||      :::|||||||||::::::   |||      |||   |||   :::
Db     681 AsnPheAspSerThrTyrGluAlaIleThrLysAspTrpPhePheAsnProThrIleThr 700

Qy    2047 CGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGC 2106
           |||   :::   |||||||||||||||:::|||:::|||:::|||:::
Db     701 ArgAlaValHisValAspIleArgTyrIleLeuTrpMetMetAspGluValAspHisAla 720

Qy    2107 CCCCGTCTCATCCAAGAGCTGAAGTTGATC 2136
                 ||||||:::||||||   |||:::
Db     721 AspAsnLeuIleLysGluLeuAspLeuVal 730
```

RESULT 7
US-10-008-355-3
; Sequence 3, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 52
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-3

```
Alignment Scores:
Pred. No.:              3.62e-15      Length:         52
Score:                  274.00        Matches:        52
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            7.17%         Indels:         0
DB:                     9             Gaps:           0

US-10-008-355-1 (1-2139) x US-10-008-355-3 (1-52)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCT 1989
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAla 20

Qy    1990 TTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGC 2049
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArg 40

Qy    2050 ACAATCAGCGTGGACATCCGCTACGTTCTCTTCATG 2085
           ||||||||||||||||||||||||||||||||||||
Db      41 ThrIleSerValAspIleArgTyrValLeuPheMet 52
```

RESULT 8
US-09-945-917-3
; Sequence 3, Application US/09945917
; Publication No. US20030042381A1
; GENERAL INFORMATION:
; APPLICANT: Bogaert, Thierry
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTIANING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/945,917
; CURRENT FILING DATE: 1998-09-21
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1528
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
US-09-945-917-3

```
Alignment Scores:
Pred. No.:              0.00144       Length:         1528
Score:                  148.50        Matches:        106
Percent Similarity:     37.48%        Conservative:   84
Best Local Similarity:  20.91%        Mismatches:     211
Query Match:            3.89%         Indels:         107
DB:                     9             Gaps:           19

US-10-008-355-1 (1-2139) x US-09-945-917-3 (1-1528)

Qy     226 ATCACAGTGTCCGATCAGGGCCTGATC--------------TTTACCAACCACCACTGC 270
           :::|||   |||:::         ||||||              :::   |||   |||
Db       1 MetThrThrSerAsnValGluLeuIleProIleTyrThrAspTrpAlaAsnArgHisLeu 20

Qy     271 GGATACGGTGCTATC---CAGAGCCAAAGCACGGTGGATCACGAGTATCTGCGCGATGGT 327
                 |||::::::   :::|||         :::   :::|||:::
Db      21 SerLysGlySerLeuSerLysSerIleArgAspIleSerAsnAspPheArgAspTyrArg 40

Qy     328 TTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTG 387
              ||||||:::   :::         :::||||||         |||      :::
Db      41 LeuValSerGlnLeuIleAsnValIleValProIleAsnGluPheSerProAlaPheThr 60

Qy     388 CGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAG 447
           :::::::::   |||:::|||         ::::::|||   |||:::         |||
Db      61 LysArgLeuAlaLysIleThrSerAsnLeuAspGly---LeuGluThrCysLeuAspTyr 79

Qy     448 ATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCA 507
           ::::::   |||                     |||   :::|||   |||   :::
Db      80 LeuLysAsnLeuGlyLeuAsp---------CysSerLysLeuThrLysThrAspIleAsp 96

Qy     508 GACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATC 567
                 |||         ::::::::         :::         :::   |||
Db      97 SerGlyAsnLeuGlyAlaValLeuGlnLeuLeuPheLeuLeuSerThrTyrLysGlnLys 116

Qy     568 GTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAG 627
           :::         :::   ||||||   :::   :::         |||:::|||:::   :::
Db     117 LeuArgGlnLeuLysLysAspGlnLysLysLeuGluGlnLeuProThrSerIle-MetPr 136
```

```
Qy    628 TTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTC 687
                                 :::
Db    136 oProAlaValSerLys------------------------------------------- 141

Qy    688 CGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTAT 747
                 :::|||   |||                 ||||||        ||||||
Db    142 -------LeuProSerProArgValAlaThrSerAlaThrAlaSerAlaThrAsnProAs 159

Qy    748 AAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATG 807
           |||   |||        :::                .:::||||||   |||
Db    159 nSerAsnPheProGlnMet-----------SerThrSerArgLeuGlnThr-------- 172

Qy    808 ACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGT 867
                   ||||||      :::|||   :::|||   :::|||:::
Db    173 ----ProGlnSerArgIleSerLysIleAspSerSerLysIleGlyIleLysProLysTh 191

Qy    868 A----------------------TCGAAAACGAGAACAATCCTCGTATCGAAGTTCGC 903
                                  |||   |||   :::        :::   |||
Db    191 rSerGlyLeuLysProProSerSerSerThrThrSerSerAsnAsnThrAsnSerPheAr 211

Qy    904 GGTATCAAGCAAGGCATCTGGA------------------------------------ 925
            ||||||::::::||||||
Db    211 gProSerSerArgSerSerGlyAsnAsnAsnValGlySerThrIleSerThrSerAlaLy 231

Qy    926 -------------------------------------------AGGAAGCCATGAGCGCA 942
                                                      :::|||   :::||
Db    231 sSerLeuGluSerSerSerThrTyrSerSerIleSerAsnLeuAsnArgProThrSerGl 251

Qy    943 GATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAG 1002
           |::::::   |||   |||            ::::::||||||   |||       ::
Db    251 nLeuGlnLysPro---SerArgProGlnThrGlnLeuValArgValAlaThrThrThrLy 270

Qy   1003 AATTCGATCG-----------GTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGT 1050
           :|||   |||                  |||||||||                 ||
Db    270 sIleGlySerSerLysLeuAlaAlaProLysAlaValSerThrProLysLeuAlaSerVa 290

Qy   1051 CGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCT 1110
           |   ::::::      |||:::                                  :::||
Db    290 lLysThrIleGlyAlaLysGlnGluProAspAsnSerGlyGlyGlyGlyGlyGlyMetLe 310

Qy   1111 GTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAAC 1170
           |   :::   ::::::         :::                   |||:::|||||
Db    310 uLysLeuLysLeuPheSerSerLysAsnProSerSerSerSerAsnSerProGlnProTh 330

Qy   1171 CGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCA 1230
           |                      |||:::::::::   ||||||:::          |||
Db    330 r---------------------ArgLysAlaAlaAlaValProGlnGlnGlnThrLeuSe 343

Qy   1231 CAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTT 1290
             :::                                     ||||||   |||
Db    343 rLysIleAlaAla----------------------------ProValLysSerGlyLeuLy 354

Qy   1291 GACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTC 1350
              ::::::|||      |||       |||      :::   |||   |||
Db    354 sProProThrSerLysLeuGlySerAlaThrSerMetSerLysLeuCysThrPro---Ly 373

Qy   1351 GATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATC 1410
           :::   |||                        |||   :::||||||   |||:::     ||
Db    373 sValSerTyrArg-----------LysThrAspAlaProIleIleSerGlnGlnAspSe 389

Qy   1411 GACAAGAAATTCAAAGGCGACACGAAGAAG-------TATGCAGACTTCGTATTCGACAA 1463
           |   |||   ||||||::::::               ||||||   |||
Db    389 rLysArgCysSerLysSerSerGluGluGluSerGlyTyrAlaGlyPhe---------As 406

Qy   1464 GAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAG---TCCATGGACAAGGA 1520
            |||      |||   |||                  |||
Db    406 nSerThrSerProThrSerSerSerThrGluGlySerLeuSerMetHisSerThrSerSe 426

Qy   1521 AAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCA---------GTAGAGCTTTCCAAGAG 1571
           |||   :::   :::   ||||||   |||:::            :::   |||:::   ||
Db    426 rLysSerSerThrSerAspGluLysSerProSerSerAspAspLeuThrLeuAsnAlaSe 446

Qy   1572 CGTAATAGCTGCTGCTCGC 1590
           |::::::   |||   |||
Db    446 rIleValThrAlaIleArg 452
```

RESULT 9
US-09-945-917-4
; Sequence 4, Application US/09945917
; Publication No. US20030042381A1
; GENERAL INFORMATION:
; APPLICANT: Bogaert, Thierry
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTIANING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/945,917
; CURRENT FILING DATE: 1998-09-21
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 1583
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
US-09-945-917-4

```
Alignment Scores:
Pred. No.:               0.00147       Length:         1583
Score:                   148.50        Matches:        106
Percent Similarity:      37.48%        Conservative:   84
Best Local Similarity:   20.91%        Mismatches:     211
Query Match:             3.89%         Indels:         107
DB:                      9             Gaps:           19

US-10-008-355-1 (1-2139) x US-09-945-917-4 (1-1583)

Qy    226 ATCACAGTGTCCGATCAGGGCCTGATC---------------TTTACCAACCACCACTGC 270
           :::|||   |||:::      ||||||               :::   |||   |||
Db      1 MetThrThrSerAsnValGluLeuIleProIleTyrThrAspTrpAlaAsnArgHisLeu 20

Qy    271 GGATACGGTGCTATC---CAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGT 327
                 |||::::::   :::|||         :::   :::|||:::
Db     21 SerLysGlySerLeuSerLysSerIleArgAspIleSerAsnAspPheArgAspTyrArg 40

Qy    328 TTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTG 387
              ||||||:::   :::         :::||||||         |||       :::
Db     41 LeuValSerGlnLeuIleAsnValIleValProIleAsnGluPheSerProAlaPheThr 60

Qy    388 CGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAG 447
           :::::::::   |||:::|||      ::::::|||   |||:::            |||
Db     61 LysArgLeuAlaLysIleThrSerAsnLeuAspGly---LeuGluThrCysLeuAspTyr 79

Qy    448 ATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCA 507
           ::::::   |||                     |||   :::|||   |||   :::
Db     80 LeuLysAsnLeuGlyLeuAsp---------CysSerLysLeuThrLysThrAspIleAsp 96

Qy    508 GACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATC 567
              |||         ::::::::         :::         :::   |||
Db     97 SerGlyAsnLeuGlyAlaValLeuGlnLeuLeuPheLeuLeuSerThrTyrLysGlnLys 116

Qy    568 GTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAG 627
           :::      :::      ||||||   :::   :::         |||:::|||::: :::
Db    117 LeuArgGlnLeuLysLysAspGlnLysLysLeuGluGlnLeuProThrSerIle-MetPr 136

Qy    628 TTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTC 687
                                 :::
Db    136 oProAlaValSerLys------------------------------------------- 141

Qy    688 CGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTAT 747
                 :::|||   |||                 ||||||         ||||||
Db    142 -------LeuProSerProArgValAlaThrSerAlaThrAlaSerAlaThrAsnProAs 159

Qy    748 AAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATG 807
           |||   |||        :::                   :::||||||   |||
```

```
Db     159 nSerAsnPheProGlnMet-----------SerThrSerArgLeuGlnThr-------- 172

Qy     808 ACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGT 867
                     ||||||     :::|||   :::|||   :::|||:::
Db     173 ----ProGlnSerArgIleSerLysIleAspSerSerLysIleGlyIleLysProLysTh 191

Qy     868 A-----------------------TCGAAAACGAGAACAATCCTCGTATCGAAGTTCGC 903
                                   |||   |||   :::          :::   |||
Db     191 rSerGlyLeuLysProProSerSerSerThrThrSerSerAsnAsnThrAsnSerPheAr 211

Qy     904 GGTATCAAGCAAGGCATCTGGA-------------------------------------- 925
                |||||:::::||||||
Db     211 gProSerSerArgSerSerGlyAsnAsnAsnValGlySerThrIleSerThrSerAlaLy 231

Qy     926 -----------------------------------------AGGAAGCCATGAGCGCA 942
                                                        :::|||   :::||
Db     231 sSerLeuGluSerSerSerThrTyrSerSerIleSerAsnLeuAsnArgProThrSerGl 251

Qy     943 GATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAG 1002
           |:::::   |||   |||                :::::||||||   |||      ::
Db     251 nLeuGlnLysPro---SerArgProGlnThrGlnLeuValArgValAlaThrThrThrLy 270

Qy    1003 AATTCGATCG-----------GTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGT 1050
           :|||   |||                      |||||||||                ||
Db     270 sIleGlySerSerLysLeuAlaAlaProLysAlaValSerThrProLysLeuAlaSerVa 290

Qy    1051 CGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCT 1110
           |   ::::::        |||:::                                :::||
Db     290 lLysThrIleGlyAlaLysGlnGluProAspAsnSerGlyGlyGlyGlyGlyGlyMetLe 310

Qy    1111 GTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAAC 1170
           |   :::   ::::::             :::                    |||:::|||||
Db     310 uLysLeuLysLeuPheSerSerLysAsnProSerSerSerSerAsnSerProGlnProTh 330

Qy    1171 CGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCA 1230
           |                     |||:::::::::   ||||||:::         |||
Db     330 r---------------------ArgLysAlaAlaAlaValProGlnGlnGlnThrLeuSe 343

Qy    1231 CAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTT 1290
              :::                                         ||||||   |||
Db     343 rLysIleAlaAla---------------------------ProValLysSerGlyLeuLy 354

Qy    1291 GACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTC 1350
                 ::::::|||      |||      |||      :::   |||   |||
Db     354 sProProThrSerLysLeuGlySerAlaThrSerMetSerLysLeuCysThrPro---Ly 373

Qy    1351 GATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATC 1410
           :::   |||                        |||   :::||||||   |||:::      ||
Db     373 sValSerTyrArg------------LysThrAspAlaProIleIleSerGlnGlnAspSe 389

Qy    1411 GACAAGAAATTCAAAGGCGACACGAAGAAG-------TATGCAGACTTCGTATTCGACAA 1463
           |   |||   ||||||::::::                    ||||||   |||
Db     389 rLysArgCysSerLysSerSerGluGluGluSerGlyTyrAlaGlyPhe---------As 406

Qy    1464 GAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAG---TCCATGGACAAGGA 1520
              |||      |||   |||                      |||
Db     406 nSerThrSerProThrSerSerSerThrGluGlySerLeuSerMetHisSerThrSerSe 426

Qy    1521 AAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCA---------GTAGAGCTTTCCAAGAG 1571
              |||   :::   :::   ||||||   |||:::            :::   |||:::   ||
Db     426 rLysSerSerThrSerAspGluLysSerProSerSerAspAspLeuThrLeuAsnAlaSe 446

Qy    1572 CGTAATAGCTGCTGCTCGC 1590
           |::::::   |||   |||
Db     446 rIleValThrAlaIleArg 452
```

RESULT 10
US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 1367
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

```
Alignment Scores:
Pred. No.:               0.0027      Length:          1367
Score:                   145.00      Matches:         160
Percent Similarity:      33.69%      Conservative:    95
Best Local Similarity:   21.14%      Mismatches:      354
Query Match:             3.80%       Indels:          148
DB:                      10          Gaps:            25

US-10-008-355-1 (1-2139) x US-09-801-368-108 (1-1367)

Qy      65 CCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAA 124
              |||   ||||||   :::      ::::::   :::|||         :::
Db     355 ProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSer 374

Qy     125 TGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCA 184
                       |||:::                         :::|||   |||:::|||   |||
Db     375 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 394

Qy     185 TTGCCAATGCCGTGGTTATCTTCGGTG------GCGGATGTACCGGTATCACAGTGTCCG 238
              :::|||   |||      |||:::               :::      ||||||:::
Db     395 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThr 414

Qy     239 ATCAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCA 298
              :::   |||   |||   |||:::::      |||         |||   |||
Db     415 ThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProVal 434

Qy     299 CGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTC 358
                    ||||||      :::                :::                |||
Db     435 ThrSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSer 454

Qy     359 CGATTCCGGGTC---TTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGG 415
                    |||      |||   |||               |||
Db     455 SerAlaProValProThrProSerSerSerThrThrGluSer------------------ 468

Qy     416 TAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGG 475
                    :::               |||   |||         |||      |||   :::
Db     469 ---SerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProValProThr 487

Qy     476 TATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGC 535
              :::   :::         :::                  |||:::|||   :::   |||
Db     488 ProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSer 507

Qy     536 CTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTA 595
              |||   |||               |||||||||:::|||         |||
Db     508 SerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAla------ 525
```

```
Qy     596 TGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGT 655
                                   ||||||      ||||||          ::::::
Db     526 --------------ProAlaProThrProSerSerSerThrThrGluSerSerSerAla 540

Qy     656 GGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGC 715
                       :::       |||   :::    ||||||      ||||||||||||
Db     541 ProValThrSerSerThrThrGluSerSerSerAla------ProValProThrProSer 558

Qy     716 CGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACT-----------TCG 763
                       |||        :::      |||::::::      |||         |||
Db     559 SerSerThrThrGluSerSerSerThrProValThrSerSerThrThrGluSerSerSer 578

Qy     764 CTGCCGTATCCA---------TGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCG 814
              |||   |||                  |||        ::::::   |||   |||:::
Db     579 AlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProValProThr 598

Qy     815 GTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAA 874
              |||   :::          :::::::|||                        |||
Db     599 ProSerSerSerThrThrGluSerSerSerAlaProAlaProThrProSerSerSerThr 618

Qy     875 ACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCA 934
            |||   :::                   :::    ::::::::::::|||
Db     619 ThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSer--------- 635

Qy     935 TGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACT 994
               |||    :::      |||   |||:::                            |||
Db     636 ---AlaProValProThrProSerSerSer-----------------------ThrThr 646

Qy     995 ATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTA 1054
                                  |||    |||      |||
Db     647 GluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAla 666

Qy    1055 AGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCC-----------GTAAGAACGGCA 1102
              ||||||            |||    |||    |||                 |||   ::::::
Db     667 ProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSer 686

Qy    1103 AGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCA 1162
                    |||    :::                |||                   |||
Db     687 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 706

Qy    1163 AGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTC 1222
              ||||||                           |||:::   ||||||
Db     707 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValPro--------- 723

Qy    1223 GTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCA 1282
                    |||      |||                          :::     |||   ||||||
Db     724 ThrProSerSerSerThrThrGluSerSerSerAlaProValProThrProSerSerSer 743

Qy    1283 AATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCG 1342
                    ::::::|||         ||||||       :::                   |||
Db     744 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 763

Qy    1343 ---------CCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATA 1393
                    |||   |||                    |||   |||        :::|||
Db     764 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValProThrProSer 783

Qy    1394 TATTCAAGAATGTAATCG------------------ACAAGAAATTCAAAGGCGACACGA 1435
              |||          |||                    |||    :::|||     |||
Db     784 SerSerThrThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGlu 803

Qy    1436 AGAAGTATG-----------CAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCG 1483
              |||:::                     |||    |||:::    :::
Db     804 SerSerValAlaProValProThrProSerSerSerSerAsnIle-------------- 818

Qy    1484 ACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGA 1543
           |||||||||    |||    |||:::|||::::::          :::
Db     819 ThrSerSerAlaProSerSerThrProPheSerSerSerThr------------------ 832

Qy    1544 AAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCG 1603

Db     833 ---------------------------------------------------GluSer 834

Qy    1604 ATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGC 1663
                    |||:::|||    |||                :::         |||
Db     835 SerSerValProValProThrProSerSerSerThrThrGluSerSerSerAlaProVal 854

Qy    1664 GTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACT---------TCACCATGC 1714
                    |||    :::                  :::||||||          |||
Db     855 SerSerSerThrThrGluSerSerValAlaProValProThrProSerSerSerSerAsn 874

Qy    1715 GTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCT------------ 1762
           :::    ::::::|||||||||                       ::::::|||
Db     875 IleThrSerSerAlaProSer-----------------SerIleProPheSerSerThr 888

Qy    1763 -----------------GGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGC 1804
                             |||||||||:::            :::      |||       |||
Db     889 ThrGluSerPheSerThrGlyThrThrValThrProSerSerSerLysTyrProGlySer 908

Qy    1805 AGGATCCTAAGAGCGATGAGT---------------------TTGCCGTACAGGAGAATA 1843
           :::          ::::::|||                        :::|||    :::
Db     909 GlnThrGluThrSerValSerSerThrThrGluThrThrIleValProThrLysThrThr 928

Qy    1844 TCCTCGACCTCTTCCGCACCAAAAACTATGGTC------------------------- 1876
           :::|||   ::::::                |||    :::
Db     929 ThrSerValThrThrProSerThrThrThrIleThrThrThrValCysSerThrGlyThr 948

Qy    1877 -------------------------GCTATGCCGAGAACGGTCAGCTCCATATCGCTT 1909
                                                |||:::||||||::::::
Db     949 AsnSerAlaGlyGluThrThrSerGlyCysSerProLysThrValThrThr--------- 965

Qy    1910 TCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACG 1969
           :::          |||||||||:::        :::||||||         :::      :::|||   |||
Db     966 ThrValProThrThrThrThrThrSerValThrThrSerSerThrThrThrIleThrThr 985

Qy    1970 GCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGT 2029
                    |||                  |||    |||                |||||| 
Db     986 ThrVal-----------CysSerThrGlyThrAsnSerAlaGlyGluThrThrSerGly 1001

Qy    2030 TCGAACCCG--------------------ATCTGCAGCGCACAATCAGCGTGGACATCC 2068
              :::|||                        ||||||                 |||
Db    1002 CysSerProLysThrIleThrThrThrValProCysSerThrSer---------ProSer 1018

Qy    2069 GCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCC 2119
              |||    |||            |||            :::    ||||||::::::
Db    1019 GluThrAlaSerGluSerThrThrThrSerProThrThrProValThrThr 1035
```

```
RESULT 11
US-09-815-242-13382
; Sequence 13382, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
```

```
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13382
;   LENGTH: 883
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-815-242-13382

Alignment Scores:
Pred. No.:              0.0225       Length:         883
Score:                  133.00       Matches:        104
Percent Similarity:     33.07%       Conservative:   65
Best Local Similarity:  20.35%       Mismatches:     166
Query Match:            3.48%        Indels:         176
DB:                     10           Gaps:           24

US-10-008-355-1 (1-2139) x US-09-815-242-13382 (1-883)

Qy    574 GATGTATTC------------------------AAGGACGTTCGTATGGTATTTGCTCCT 609
          |||||||||                        |||       :::|||     |||
Db     26 AspValPheLysProSerGlyAspGlnLysAlaLysProTyrSerIleValIleProPro 45

Qy    610 CCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACG 669
             |||:::     ||||||                                  |||
Db     46 ProAsnValThrGlyLysLeu--------------------------------HisLeu 54

Qy    670 GGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGC 729
          |||                                                  ::::::
Db     55 Gly---------------------------------------------------HisAla 57

Qy    730 AAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAG 789
             |||          :::   :::                  |||||||||:::
Db     58 TrpAspThr---ThrLeuGlnAspIleIleIleArgGlnLysArgMetGlnGlyPheAsp 76

Qy    790 GCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTC---ACT 846
                                |||:::     ||||||           :::    |||
Db     77 ------------------ThrLeuTrpLeuProGlyMetAspHisAlaGlyIleAlaThr 90

Qy    847 TCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGT 906
                   ||||||:::|||:::      |||        |||    ::::::
Db     91 GlnAlaLysValGluGluArgLeuArgGlyGluGlyIleSerArgTyrAspLeuGlyArg 110

Qy    907 ------------ATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACC 954
                      |||        ||||||:::               :::   ||||||
Db    111 GluSerPheLeuThrLysValTrpGluTrpLysAsp------------GluTyrAlaThr 126

Qy    955 CGTATCAAATATGCC---AGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATC 1011
             ||||||         |||          |||    :::|||     :::
Db    127 ThrIleLysGluGlnTrpGlyLysMetGlyLeuSerValAspTyrSerArgGluArgPhe 146

Qy   1012 GGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGA 1071
             ::::::   ||||||:::                        :::|||        :::
Db    147 ThrLeuAspGluGlyLeuSer-----------------------LysAlaValArgLys 158

Qy   1072 GCATTCGCAGACTGGATCCGTAAG------------AACGGCAAGAGTGCTGTCTATGGC 1119
             |||   |||        :::|||                |||:::        :::
Db    159 ValPheValAspLeuTyrLysLysGlyTrpIleTyrArgGlyGluPheIleIleAsnTrp 178

Qy   1120 GAT---------------GTATTGTCTTCTCTCGAAAAGGCTTATAAG------------ 1152
          |||                            ||||||   :::|||      :::|||
Db    179 AspProAlaAlaArgThrAlaLeuSerAspIleGluValIleHisLysAspValGluGly 198

Qy   1153 ------------------------------GAAGGAGCCAAGGCCAACCGTGAGATGACT 1182
                                        :::|||::::::|||            |||
Db    199 AlaPheTyrHisMetAsnTyrMetLeuGluAspGlySerArgAlaLeuGluValAlaThr 218

Qy   1183 TATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAAC 1242
                   ||||||:::||||||
Db    219 ThrArgProGluThrMetPheGlyAspVal----------------------------- 228

Qy   1243 GCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTAC 1302
          |||:::|||    ||||||:::                             |||    :::|||
Db    229 AlaIleAlaValAsnProGlu--------------------------AspProArgTyr 239

Qy   1303 AAAGACTACCTCCCCTCG---------------CTCGACCGTAAGGTGCTGCCCGCCATG 1347
          ||||||    :::                    :::    |||::::::|||    :::
Db    240 LysAspLeuIleGlyLysAsnValIleLeuProIleAlaAsnLysLeuIleProIleVal 259

Qy   1348 CTCGAT--------------------------------------ATTGTACGCCGGCGT 1368
             |||                                       |||
Db    260 GlyAspGluHisAlaAspProGluPheGlyThrGlyValValLysIleThrProAlaHis 279

Qy   1369 ATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC 1428
             |||   |||   |||                  |||:::
Db    280 AspProAsnAspPheLeuValGlyGlnArgHisAsnLeuProGlnValAsnIleMetAsn 299

Qy   1429 GACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAG 1488
          |||               :::   ||||||:::   |||    :::            |||:::
Db    300 AspAspGlyThrMetAsnGluLeuValPheGluPheSerGlyMet---------AspArg 316

Qy   1489 TTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGAT 1548
          |||   |||                  |||      ||||||      |||
Db    317 PheGluAlaArg---------------LysAlaValValAlaLysLeuGluGluIleGly 331

Qy   1549 CCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCT--------------- 1593
             |||::::::   |||   |||   :::            :::
Db    332 AlaLeuValLysIleGluLysArgValHisSerValGlyHisSerGluArgThrGlyVal 351

Qy   1594 ------------------------------------------------ATTCAGGCCGAT 1605
                                                          :::   :::
Db    352 ValValGluProArgLeuSerThrGlnTrpPheValLysMetAspGlnLeuAlaLysAsn 371

Qy   1606 GCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGT 1665
          |||:::||||||          :::|||
Db    372 AlaIleAlaAsnGlnAspThrGluAspLys-------------------------Val 382

Qy   1666 GAGATGTACCCCGGACGTGCTCTGCCGAGC--------------------------GAT 1698
          |||   ||||||   |||          :::                         |||
Db    383 GluPheTyrProProArgPheAsnAspThrPheLeuGlnTrpMetGluAsnValHisAsp 402

Qy   1699 GCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGT 1758
                :::   ::::::   :::|||                  |||
Db    403 TrpValIleSerArgGlnLeuTrpTrpGlyHis------------------GlnIlePro 416

Qy   1759 GCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGC 1818
          ||||||||||||          |||:::      :::   ::::::   |||:::
Db    417 AlaTrpTyrAsn------AlaAspGlyGluMetTyrValGlyGluGluAlaProGluGly 434

Qy   1819 GATGAGTTTGCCGTACAGGAGAATATCCTCGAC 1851
          |||   :::         |||::::::||||||
Db    435 AspGlyTrpThrGlnAspGluAspValLeuAsp 445

RESULT 12
US-09-815-242-13684
; Sequence 13684, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
```

```
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13684
;   LENGTH: 883
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-815-242-13684

Alignment Scores:
Pred. No.:                0.0735      Length:          883
Score:                    127.00      Matches:         103
Percent Similarity:       32.88%      Conservative:    65
Best Local Similarity:    20.16%      Mismatches:      167
Query Match:              3.32%       Indels:          176
DB:                       10          Gaps:            24

US-10-008-355-1 (1-2139) x US-09-815-242-13684 (1-883)

Qy    574 GATGTATTC------------------------AAGGACGTTCGTATGGTATTTGCTCCT 609
          |||||||||                        |||        :::|||       |||
Db     26 AspValPheLysProSerGlyAspGlnLysAlaLysProTyrSerIleValIleProPro 45

Qy    610 CCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACG 669
          |||:::      ||||||                                     |||
Db     46 ProAsnValThrGlyLysLeu---------------------------------HisLeu 54

Qy    670 GGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGC 729
          |||                                                   ::::::
Db     55 Gly---------------------------------------------------HisAla 57

Qy    730 AAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAG 789
             |||         :::    :::                  |||||||||:::
Db     58 TrpAspThr---ThrLeuGlnAspIleIleIleArgGlnLysArgMetGlnGlyPheAsp 76

Qy    790 GCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTC---ACT 846
                            |||:::      ||||||            :::    |||
Db     77 ------------------ThrLeuTrpLeuProGlyMetAspHisAlaGlyIleAlaThr 90

Qy    847 TCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGT 906
                   |||||||:::|||:::      |||          |||    ::::::
Db     91 GlnAlaLysValGluGluArgLeuArgGlyGluGlyIleThrArgTyrAspLeuGlyArg 110

Qy    907 ------------ATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACC 954
                         |||      ||||||:::               :::    ||||||
Db    111 GluSerPheLeuThrLysValTrpGluTrpLysAsp------------GluTyrAlaThr 126

Qy    955 CGTATCAAATATGCC---AGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATC 1011
             ||||||           |||          |||    :::|||    :::
Db    127 ThrIleLysGluGlnTrpGlyLysMetGlyLeuSerValAspTyrSerArgGluArgPhe 146

Qy   1012 GGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGA 1071
             ::::::   |||||||:::                        :::|||      :::
Db    147 ThrLeuAspGluGlyLeuSer------------------------LysAlaValArgLys 158

Qy   1072 GCATTCGCAGACTGGATCCGTAAG------------AACGGCAAGAGTGCTGTCTATGGC 1119
             |||   |||       :::|||               |||:::      :::
Db    159 ValPheValAspLeuTyrLysLysGlyTrpIleTyrArgGlyGluPheIleIleAsnTrp 178

Qy   1120 GAT---------------GTATTGTCTTCTCTCGAAAAGGCTTATAAG------------ 1152
          |||               ||||||   :::|||      :::|||
Db    179 AspProAlaAlaArgThrAlaLeuSerAspIleGluValIleHisLysAspValGluGly 198

Qy   1153 ------------------------------GAAGGAGCCAAGGCCAACCGTGAGATGACT 1182
                                        :::|||::::::|||          |||
Db    199 AlaPheTyrHisMetAsnTyrMetLeuGluAspGlySerArgAlaLeuGluValAlaThr 218

Qy   1183 TATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAAC 1242
                     ||||||:::||||||
Db    219 ThrArgProGluThrMetPheGlyAspVal------------------------------ 228

Qy   1243 GCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTAC 1302
          |||:::|||   ||||||:::                           |||   :::|||
Db    229 AlaValAlaValAsnProGlu---------------------------AspProArgTyr 239

Qy   1303 AAAGACTACCTCCCCTCG---------------CTCGACCGTAAGGTGCTGCCCGCCATG 1347
          ||||||   :::                     :::   |||::::::|||   :::
Db    240 LysAspLeuIleGlyLysAsnValIleLeuProIleAlaAsnLysLeuIleProIleVal 259

Qy   1348 CTCGAT---------------------------------------ATTGTACGCCGGCGT 1368
             |||                                       |||
Db    260 GlyAspGluHisAlaAspProGluPheGlyThrGlyValValLysIleThrProAlaHis 279

Qy   1369 ATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC 1428
             |||   |||   |||               |||:::
Db    280 AspProAsnAspPheLeuValGlyGlnArgHisAsnLeuProGlnValAsnValMetAsn 299

Qy   1429 GACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAG 1488
          |||          :::   |||:::   |||   :::            |||:::
Db    300 AspAspGlyThrMetAsnGluLeuAlaPheGluPheSerGlyMet---------AspArg 316

Qy   1489 TTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGAT 1548
          |||   |||               |||         ||||||       |||
Db    317 PheGluAlaArg---------------LysAlaValValAlaLysLeuGluGluIleGly 331

Qy   1549 CCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCT--------------- 1593
                |||::::::   |||   |||   :::         :::
Db    332 AlaLeuValLysIleGluLysArgValHisSerValGlyHisSerGluArgThrGlyVal 351

Qy   1594 ---------------------------------------------ATTCAGGCCGAT 1605
                                                       :::   :::
Db    352 ValValGluProArgLeuSerThrGlnTrpPheValLysMetAspGlnLeuAlaLysAsn 371

Qy   1606 GCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGT 1665
          |||:::||||||      :::|||
Db    372 AlaIleAlaAsnGlnAspThrGluAspLys---------------------------Val 382

Qy   1666 GAGATGTACCCCGGACGTGCTCTGCCGAGC---------------------------GAT 1698
          |||   ||||||   |||         :::                           |||
Db    383 GluPheTyrProProArgPheAsnAspThrPheLeuGlnTrpMetGluAsnValHisAsp 402

Qy   1699 GCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGT 1758
             :::   ::::::   :::|||                              |||
Db    403 TrpValIleSerArgGlnLeuTrpTrpGlyHis------------------GlnIlePro 416

Qy   1759 GCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGC 1818
          ||||||||||||         |||:::      :::   ::::::   |||:::
Db    417 AlaTrpTyrAsn------AlaAspGlyGluMetTyrValGlyGluGluAlaProGluGly 434

Qy   1819 GATGAGTTTGCCGTACAGGAGAATATCCTCGAC 1851
          |||   :::         |||::::::||||||
Db    435 AspGlyTrpThrGlnAspGluAspValLeuAsp 445

RESULT 13
US-10-008-355-18
; Sequence 18, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
```

```
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
;   LENGTH: 26
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-18

Alignment Scores:
Pred. No.:                0.0237        Length:          26
Score:                    122.50        Matches:         25
Percent Similarity:       96.15%        Conservative:    0
Best Local Similarity:    96.15%        Mismatches:      0
Query Match:              3.21%         Indels:          1
DB:                       9             Gaps:            1

US-10-008-355-1 (1-2139) x US-10-008-355-18 (1-26)

Qy      70 GCCGACAAAGGC---ATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATG 126
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db       1 AlaAspLysGlyMetMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArgMet 20

Qy     127 CGTGAGCTCGGCTTTACG 144
           ||||||||||||||||||
Db      21 ArgGluLeuGlyPheThr 26

RESULT 14
US-10-192-584-6
; Sequence 6, Application US/10192584
; Publication No. US20030027987A1
;   GENERAL INFORMATION:
;        APPLICANT: TOKUNAGA, Eiji
;                   SAKAGUCHI, Masashi
;                   MATSUO, Kazuo
;                   HAMADA, Fukusaburo
;                   TOKIYOSHI, Sachio
;        TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
;                            PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
;        NUMBER OF SEQUENCES: 8
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK
;             STREET: 624 Ninth Street, N.W., Suite 300
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: USA
;             ZIP: 20001
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Patent In Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/192,584
;             FILING DATE: 11-Jul-2002
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/077,098
;             FILING DATE: 19-May-1998
;             APPLICATION NUMBER: PCT/JP97/03222
;             FILING DATE: 12-SEP-1997
;             APPLICATION NUMBER: JP 27,148/1996
;             FILING DATE: 19-SEP-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: KORNBAU, Anne M.
;             REGISTRATION NUMBER: 25,618
;             REFERENCE/DOCKET NUMBER: TOKUNAGA=1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-628-5197
;             TELEFAX: 202-737-3528
;   INFORMATION FOR SEQ ID NO: 6:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2042 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-192-584-6

Alignment Scores:
Pred. No.:                0.636         Length:          2042
Score:                    118.50        Matches:         140
Percent Similarity:       36.03%        Conservative:    87
Best Local Similarity:    22.22%        Mismatches:      223
Query Match:              3.10%         Indels:          180
DB:                       9             Gaps:            35

US-10-008-355-1 (1-2139) x US-10-192-584-6 (1-2042)

Qy     550 AACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCT 609
           |||:::   |||   |||   |||   |||   |||         :::|||
Db       2 AsnLysValPheLysIleLysTyrSerValValLysGlnGluMetIleVal--------- 18

Qy     610 CCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACG 669
                    |||   :::         :::   |||            ::::::|||
Db      19 ---------ValSerGluLeuAlaAsnAsnLysAspLysThrAlaSerGlnLysAsnThr 35

Qy     670 GGCGACTTCAGCGTATTCCGC--------------------------------------- 690
              :::   :::   |||:::
Db      36 HisAsnThrAlaPhePheGlnProLeuPheThrLysCysThrTyrLeuAlaLeuLeuIle 55

Qy     691 ---GTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTAT 747
              :::      ||||||         |||            |||      :::
Db      56 AsnIleAlaLeuGlyAlaSerLeuPhePro---------GlnLeuAlaAsnAlaLysTrp 72

Qy     748 AAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGAC--------- 798
                 ||||||   ::::::|||   :::               |||
Db      73 LeuGluValTyrSerSerSerValLysLeuSerThrValSerAlaGlnSerAsnSerVal 92

Qy     799 ---------------------------------------------TATGCCATGACCATC 813
                                                           :::   |||
Db      93 AsnLeuAsnProSerGlyAlaGluSerValGlyThrAsnSerProGlnGlyValAlaIle 112

Qy     814 GGTTTCCCGGGCAGTACGGATCGCTACCTCACT---------TCTTGGGGTGTGGAAGAT 864
           |||:::      :::   ||||||         |||         :::   ||||||
Db     113 GlyTyrGlyAlaThrAsnAspArgSerAlaThrGlyAlaIleAlaLeuGlyVal------ 130

Qy     865 CGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGG 924
              ::::::||||||         :::         |||      |||
Db     131 GlyValLysAsnGluThrLeuAlaLysAspSerIle---AlaIleGlyTyrGlyAlaLys 149

Qy     925 AAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAG 984
              |||:::   :::               |||      |||                  |||
Db     150 AsnGluSerThrAlaProSerSerValThrIleGlyLys-----------------Gln 163

Qy     985 AGTGCTAACTATTGGAAGAATTCGATC------GGTATGAACCGCGGTCTCGCTCGTCTT 1038
           :::   |||   ::::::   ||||||         |||:::|||         :::|||
Db     164 AlaIleAsnArgPheGluLysSerIleValMetGlyLeuAsn---AlaTyrThrGlnLeu 182

Qy    1039 GACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAAC 1098
           |||      |||         :::   |||   |||         :::      |||   :::|||
Db     183 AspProArgGlyThrSerLys---GluThrArgGlnGlySerValValIleGlyGluAsn 201

Qy    1099 GGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGA 1158
              |||||||||         |||:::         |||:::
Db     202 AlaLysSerAla------GlyAsn------GlnSerValSerLeuGlyGlnAsnSerTrp 217

Qy    1159 GCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTG 1218
           :::|||   |||                                 |||   ::::::      ||||||
Db     218 SerLysThrAsn-----------------SerIleSerIleGlyAlaGlyThr------ 229

Qy    1219 GTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATC 1278
                 ||||||:::         :::::::::|||         |||      :::|||
Db     230 ------PheAlaGluGlyLysSerSerIleAlaIleGlyThrAspLysIleSerGly--- 246

Qy    1279 CTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTG 1338
                                 ||||||   |||   ||||||:::
```

```
Db      247 --------------ThrLysTyrAsnAspLysLeuProAlaThrAlaTrpAsnGlyThr 261

Qy     1339 CCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTC 1398
                                          :::|||   :::   :::   |||||||||
Db      262 GlyThr-----------------------ValProLysAsnSerIleTrpAspIlePhe 273

Qy     1399 AAGAATGTA---ATCGACAAGAAATTCAAAGGC---------GACACGAAG--------- 1437
                  :::   :::   |||:::        |||         ||||||
Db      274 SerGluLeuTyrMetGlyLysGlnThrAsnGlyArgAspTyrAspThrThrThrArgAsp 293

Qy     1438 -----------------------AAGTATGCAGACTTC-----------GTATTCGAC 1461
                                   |||::::::||||||              |||
Db      294 ProAsnLysProGluAlaPheTyrLysPheSerAspPheLysGlyLysTyrValAsnThr 313

Qy     1462 AAGAGTGTGGTTCCT---TATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAG 1518
               :::      |||   |||:::   |||      |||:::
Db      314 ProThrAlaSerProThrTyrAlaGlyLysLeuGlyAlaIle------------------ 327

Qy     1519 GAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTA--- 1575
                     |||                    |||   |||:::|||    :::|||
Db      328 ---------AlaLeuGlySerArgThrIleAlaAlaGlyGluMetSerThrAlaValGly 344

Qy     1576 ---ATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAG 1632
               :::|||   |||   |||   ::::::   ||||||
Db      345 SerLeuAlaPheAlaLeuAlaAspArgSerThrAlaMet--------------------- 357

Qy     1633 AAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCG 1692
               |||   |||   |||   |||         ::::::         |||   |||:::
Db      358 ---GlyLeuArgSerPheValAla------LysAspAlaValGlyGlyThrAlaIleGly 374

Qy     1693 AGCGATGCC--------------------------------AACTTCACCATGCGTATG 1719
            ::::::                                   |||   |||
Db      375 GluGluSerArgThrPheAlaLysAspSerValAlaIleGlyAsnLysThrGluAlaSer 394

Qy     1720 AGCTACGGCTCCATCAAG------------------------------------------ 1737
            :::   ||||||:::
Db      395 AsnAlaGlySerMetAlaTyrGlyTyrLysAlaLysAlaValGlyAlaGlyAlaIleAla 414

Qy     1738 ---GGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTA 1794
               |||   |||         ||||||   :::|||   |||   ||||||      :::
Db      415 IleGlyThrGluValAlaAlaGlyAlaLysPheAsnSerHisGlnThrGlyAsnLeuLeu 434

Qy     1795 TTGGAG--------------------AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTA 1833
            :::                       |||      |||||||||:::
Db      435 GlnAspAsnAsnAlaTyrAlaThrLeuLysAsnAlaAspLysSerAspAsp---ThrLys 453

Qy     1834 CAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGT 1893
                  |||   :::   :::|||   |||:::::::::                ||||||
Db      454 ThrGlyAsnAlaIleThrValPhe---ThrGlnSerPheAspAsnMetLeuThrAsnGly 472

Qy     1894 CAGCTCCATATCGCTTTCCTATCGAACAACGACATC----------------------- 1929
                     :::      :::|||   |||:::
Db      473 ---------LeuProLeuValSerGluAsnGluThrTyrLeuThrThrSerAlaGlyAla 489

Qy     1930 ---------ACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
                     |||            |||         ||||||            |||
Db      490 IleLysLysThrAlaThrThrAspSerSerAlaGlyGlyGlyLysAsnAlaIleAlaIle 509

Qy     1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
            |||      :::   |||:::   |||      |||         |||         ::
Db      510 GlySerLys-ThrPheAlaSer---LysAlaAsnSerValAlaLeuGlySerTyrAlaLe 528

Qy     2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCT---------TCATGATTGAC 2091
            :    ||||||:::|||:::        ::::::|||         |||      ||
Db      528 uAlaAspAlaGlnAsnAlaPheAlaLeuGlySerTyrSerPheValGluSerSerAlaTh 548

Qy     2092 AAATGGGGTCAGTGCCCCCGTCTCATCC 2119
             ||||   ::::::  .   |||   |||
Db      548 rAsnThrIleThrIleGlyValGlySer 557
```

RESULT 15
US-09-920-552-103
; Sequence 103, Application US/09920552
; Patent No. US20020094576A1
; GENERAL INFORMATION:
; APPLICANT: Griffiths, David J.
; APPLICANT: Weiss, Robin A.
; APPLICANT: Venables, Patrick
; TITLE OF INVENTION: Material and Methods Relating to a No. US20020094576A1el Retro
; FILE REFERENCE: Abbott Labs
; CURRENT APPLICATION NUMBER: US/09/920,552
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 09/280,329
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: GB 9806649.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/115,288
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 103
; LENGTH: 1814
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-920-552-103

```
Alignment Scores:
Pred. No.:               0.656      Length:         1814
Score:                   118.00     Matches:        162
Percent Similarity:      33.29%     Conservative:   89
Best Local Similarity:   21.49%     Mismatches:     252
Query Match:             3.09%      Indels:         251
DB:                      10         Gaps:           43

US-10-008-355-1 (1-2139) x US-09-920-552-103 (1-1814)

Qy      255 TACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTA 314
               |||   ||||||   |||                  |||   ||||||   ::::::
Db       47 TyrSerProProProArgSerPheSerSerLysArgGlyLysArgGlyGlyArgLysIle 66

Qy      315 TCTGCG-------------------------CGATGGTTT------CGTTTCTCGCAC 341
               |||                        ||||||         :::::: |||
Db       67 HisAlaPheValArgAlaValValGluMetSerProTrpPheLeuGlnGlnTrp-ArgGl 86

Qy      342 GATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAA 401
               :::|||   :::|||         |||   |||::::::   |||   |||:::   ::
Db       86 nValGlyTrpLysLeuArgGluAlaGlyGluSerIleGluValLeu---LysLeuTrpGl 105

Qy      402 GGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAG-------- 453
            ::::   ::::||                   :::   |||   ||||||
Db      105 nLeuValAspGlnVal-----------------LeuGlnAspProMetGluGlyIleGl 119

Qy      454 ----CGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGA 509
                  |||   :::      |||:::   :::||||||         :::    :::
Db      119 uLeuLeuLeuAsnGluGlySerGluLeuGlnGluGluLeu------TrpGlnGlySerGl 137

Qy      510 CGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGT 569
             |||:::      |||      |||               ::::::      |||
Db      137 nGluSerSerTyrCys---AsnGluGlnValArgArgLysSerLysLysAspLeuGluSe 156

Qy      570 CTACGATGTATTC---------AAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGT 620
               :::   :::            |||:::               :::||||||      ::
Db      156 rGlnGluAlaTyrValGlyGluLysGluAlaGluGluGlyThrSerProProArgGluMe 176

Qy      621 AGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAG 680
            :
Db      176 tPro--------------------------------------------------------- 177

Qy      681 CGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAA 740
                      |||:::|||   |||                     ||||||
Db      178 -------ArgIleTyrProGlyProThrArgLeuAlaAsnSerThrIleLysAspGluSe 195

Qy      741 ACCC------TATAAGCCCGTTTAC---------TTCGCTGCCGTATCCATGCAAGGCTA 785
             |||       |||:::|||:::          |||      |||               ::
Db      195 rProProArgTyrArgProLeuValLeuProLysPheProProValLysLysSerProPh 215
```

```
Qy     786 CAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCAC 845
           :                                 ||||||      :::
Db     215 e----------------------------PheProHisGluIleGluGlu-------- 222

Qy     846 TTCTTGG----------------------------GGTGTGGAAGATCGTATCGAAAA 875
              |||                             |||:::::::::  ||||||
Db     223 -GluTrpGluGluGluLysGlyValMetGluGlyValGlyMetGlnGluGluIleGlu-- 241

Qy     876 CGAGAACAATCCTCGTATCGAA------GTTCGCGGT------ATCAAGCAAGGCATCTG 923
                  ||||||  |||         ||||||       ::::::    ||
Db     242 -------ArgProArgThrGluSerLysGlnArgGlyAlaValTyrArgLysProAlaTr 259

Qy     924 GAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC---------------------- 960
           |::::::              |||        :::
Db     259 pGluGlnTyrGlnGluProGlyGlnValPheProValIleGlnAspGlnGlnGluAsnAr 279

Qy     961 ---------------------AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTG 998
                                |||        :::    ||||||           ::
Db     279 gGlyTrpThrProLeuAspHisLysLeuLeuLysGluLeuGlnGlnSerValGlnLeuTy 299

Qy     999 GAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCG 1058
           :                          :::   :::       ||||||  ||||||:::::
Db     299 rGlyProHisAlaSerTyrThrGlnAlaIle------LeuAspAsnIleGlyGlnGlnGl 317

Qy    1059 TGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTAT-- 1116
                                ||||||     :::|||        ||||||
Db     317 yLeuIleProGlu---------AspTrp------ArgAsnLeuValLysAlaValLeuGl 332

Qy    1117 ----GGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCG 1172
                ||||||   |||        |||      ||||||||||||    |||:::
Db     332 yGlyGlyAspPheLeuLeuTrpLeu---AlaAlaTyrLysGluPheAlaArgGluTyrSe 351

Qy    1173 TGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTT---------- 1221
           :::      |||     :::                  |||::::::
Db     351 rLysGlnAsnTyrGlnAlaGlyAsnGlnAlaTrpAspGluGluMetLeuIleGlyGluGl 371

Qy    1222 -CGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCT 1280
             |||||||||                       |||||||||    |||
Db     371 yArgPheAla-----------------------AsnProAspGluGlnAla-------- 380

Qy    1281 CAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCC 1340
                                                  |||        ||||||
Db     381 -----------------------------------ArgLeuProThrAlaValLeuLe 388

Qy    1341 CGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAA 1400
                 :::   :::|||  ||||||                                  ::
Db     388 uGlnValArgGluIleAlaArgArgAlaTrpLysValIleProArgLysGlyGluIleAr 408

Qy    1401 GAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTA----- 1455
            ::::   :::    |||       :::|||   |||:::   ||||||||||||||
Db     408 gHisSerLeuThrLysIleValGlnGlyProThrGluProTyrAlaAspPheValAsnAr 428

Qy    1456 -----------------------TTCGAC------AAGAGTGTGGTTCCT-------- 1476
                                  |||:::      :::       :::|||
Db     428 gLeuMetGluAlaAlaGlyAsnIlePheGluThrValGluGluAlaMetProLeuValAr 448

Qy    1477 ---------TATAGCGACAAGTTCCATGCCATGCTCAAGT---------CCATGGACAA 1517
                    |||       ::::::        |||             ||||||
Db     448 gArgLeuAlaTyrGluGln-AlaAsnLysThrCysArgGluAlaLeuArgProTrpGlnH 468

Qy    1518 GGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAG---------------TAGAGCT 1562
              |||   :::|||         :::      |||                  |||
Db     468 isLysAspIleProThrPheLeuLysIleCysArgAspValMetAspAspValAlaSerG 488

Qy    1563 TTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTA 1622
              |||                :::|||:::   |||       |||        |||
Db     488 lySerHisAla---------AlaValAlaValThrArgGlnPheArgGlnAsnMet---- 503

Qy    1623 TGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACG 1682
                   ||| |||:::      |||    |||                 :::    |||
Db     504 --------Arg-GlyArgIleCysPheArgCysGlyArgGluGlyHisLeuMetArgGlu 520

Qy    1683 TGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATA 1742
           |||   |||:::                     :::     |||       |||||||||
Db     521 CysArgAlaLys--------------------PheProAlaArgAspArgGlnGlyIle 533

Qy    1743 TGAACCGCA-----------------GGACGGTGCCTGGTACAACTATCATACGACAGG 1784
                 :::                 ||||||                         |||
Db     534 GluAsnSerArgArgProGlyLeuCysProArgCys--------------------Arg 546

Qy    1785 CAAGGG-------------------------CGTATTGGAGAAGCAGGATCCTAAGAG 1817
           :::|||                         :::|||      ||||||:::     :::
Db     547 ArgGlyAsnHisTrpValSerGluCysTyrPheLysIleAsnSerAlaGlyAsnProGln 566

Qy    1818 CGA-----------------TGAGTTTGCCGTACAGGAGAATAT--------------- 1844
           |||                 :::    |||:::|||     |||
Db     567 ArgMetProAsnGluProGlyProIleArgArgSerGlyValTyrGlyGlnGlnGlnLys 586

Qy    1845 --------------CCTCGACCTCTT------------------------------- 1856
                         |||   |||
Db     587 AsnGlyMetTrpAlaProAlaProArgGlyGlnAsnAsnArgTrpValProArgSerGln 606

Qy    1857 ---------------------------------------------------CCGCAC 1862
                                                              ||||||
Db     607 AsnGlnIlePheThrAlaGlnSerTyrGluProProAsnIleArgTyrAlaGlyProHis 626

Qy    1863 CAAAAA---------CTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCT 1913
           :::|||         :::||||||      ::::::|||:::                |||
Db     627 LysLysIleArgCysIleTrpSerThrAlaLysLysArgAsnValGlyProCysThrPro 646

Qy    1914 ATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATT---CGATAAGAA--- 1967
              ::::::      |||      :::       |||   |||   |||   |||
Db     647 GlyProLysGlnSerValGlyAlaThrValProLysProAsnIleTyrArgAlaGluLeu 666

Qy    1968 CGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGA 2027
           |||   ::::::   |||             |||:::|||                 |||
Db     667 ArgAlaThrGluTyrSerValLeuThrProGlnMetGlyProGlnAlaIleGluValArg 686

Qy    2028 GTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGA 2063
                 |||   ||||||   ||||||      |||
Db     687 AlaGluThrIleSerAlaGluHisAsnGlyLeuGly 698
```

Search completed: May 16, 2003, 13:03:11
Job time : 124.5 secs

```
OM nucleic - protein search, using frame_plus_n2p model

Run on:          May 16, 2003, 12:22:51 ; Search time 79.5 Seconds
                                          (without alignments)
                                          5173.120 Million cell updates/sec

Title:           US-10-008-355-1
Perfect score:   3820
Sequence:        1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters:      566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110338_9364/app_query.fasta_1.2311
-DB=PIR_73 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_111_@runat_16052003_110338_9364 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       PIR_73:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	907.5	23.8	716	2	G82627	hypothetical prote
2	156	4.1	1791	2	T02345	hypothetical prote
3	154.5	4.0	1609	2	S25345	probable membrane
4	149	3.9	528	2	I47141	gastric mucin (clo
5	145	3.8	1367	1	S48478	glucan 1,4-alpha-g
6	143.5	3.8	776	2	S44784	C30C11.4 protein -
7	141	3.7	948	2	T11678	hypothetical prote
8	139.5	3.7	1802	2	S69703	HKR1 protein precu
9	139	3.6	4543	1	A53102	alpha-2-macroglobu
10	138	3.6	477	2	S53362	mucin 5AC (clone J
11	137	3.6	3570	2	T45025	mucin MUC5B, trach
12	133.5	3.5	1104	2	A60999	alpha-amylase (EC
13	133	3.5	883	2	D97933	valine-tRNA ligase
c 14	132	3.3	571	2	C71528	probable pts pep p
15	131	3.4	648	2	PC4395	mucin 3 - human (f
16	129.5	3.4	1104	2	S59310	probable membrane
17	129	3.4	1104	1	A36866	microbial collagen
18	127.5	3.3	725	2	A41258	a-agglutinin core
19	127	3.3	2232	2	T34434	hypothetical prote
20	126.5	3.3	666	1	D69103	DNA helicase (EC 3
21	126.5	3.3	796	2	T21460	hypothetical prote
22	125	3.3	1440	2	T33813	hypothetical prote
23	124.5	3.3	735	2	F98228	1,4-alpha-glucan b
24	124.5	3.3	735	2	AH3057	glycogen branching
25	124.5	3.3	799	2	T25174	hypothetical prote
26	123	3.2	1289	2	F72308	hypothetical prote
27	123	3.2	1469	2	T19168	hypothetical prote
28	123	3.2	1530	2	H83865	glutamate synthase
29	122.5	3.2	711	2	S70660	transferrin-bindin
30	122.5	3.2	1571	2	AC1647	Lactobacillus phag
31	122	3.2	721	2	AD1617	penicillin-binding
32	122	3.2	743	2	D82883	DNA helicase II UU
33	122	3.2	883	2	C95066	valyl-tRNA synthet
34	120.5	3.2	292	2	S24169	mucin - rat
35	120.5	3.2	779	2	T25173	hypothetical prote
c 36	120.5	3.0	1791	2	T02345	hypothetical prote
37	120.5	3.2	3020	2	A43932	mucin 2 precursor,
38	120	3.1	1275	2	T33369	hypothetical prote
39	119	3.1	529	2	T51214	related to trichod
40	119	3.1	825	2	T29634	hypothetical prote
41	119	3.1	1167	2	AF0258	probable phage tai
42	118.5	3.1	1777	2	T34369	hypothetical prote
43	118	3.1	698	2	C82332	translation elonga
c 44	117.5	3.0	1122	2	T14180	exiT protein - Myc
45	117	3.1	1147	2	T40866	cell polarity prot

ALIGNMENTS

```
RESULT 1
G82627
hypothetical protein XF1887 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: G82627
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: G82627
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-716 <SIM>
A;Cross-references: GB:AE004008; GB:AE003849; NID:g9106961; PIDN:AAF84693.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1887

Alignment Scores:
Pred. No.:            7.33e-54       Length:         716
Score:                907.50         Matches:        239
Percent Similarity:   49.21%         Conservative:   135
```

```
Best Local Similarity:  31.45%       Mismatches:    291
Query Match:            23.76%       Indels:        95
DB:                     2            Gaps:          21

US-10-008-355-1 (1-2139) x G82627 (1-716)

Qy      1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
          |||:::      |||   |||:::    :::  |||   |||:::         |||
Db      1 MetArgPheAsnLeuLeuSerLeuSerValLeuAlaThrLeuIleThrValAspSer--- 19

Qy     61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
             |||   :::||||||||||:::       :::|||     |||
Db     20 ---ThrHisAlaGlyGluGlyMetTrpValProGlnGlnLeu---ProGluIleAlaGly 37

Qy    121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
             :::::::::   |||      |||          :::   |||   :::
Db     38 ProLeuLysGlnAlaGlyLeuGlnLeuSerProGluGlnLeuSerAsnLeuThrGlyAsp 57

Qy    181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
             :::          |||   :::       |||   ||||||         ||||||
Db     58 ProMetGlyAlaValValSerLeu------GlyAsnCysThrAlaSerLeuValSerPro 75

Qy    241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
          :::||||||:::   ||||||||||||||   ||||||||||||||         ||||||
Db     76 GluGlyLeuValIleThrAsnHisHisCysAlaTyrGlyAlaIleGlnLeuAsnSerThr 95

Qy    301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
                :::   :::::::::||||||   :::   |||      :::|||:::
Db     96 ProLysLysAsnLeuIleLysGluGlyPheAsnAlaLeuThrGlnAlaAspGluValSer 115

Qy    361 ATT---CCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTA 417
             |||         :::      |||   :::|||      ||||||   :::
Db    116 AlaGlyProAsnAlaArgIleTyrValLeuGluGlnIleThrAspValThrAlaGlnAla 135

Qy    418 GAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTA 477
          :::      |||         ::::::      :::         |||   |||
Db    136 LysAlaAlaLeuAlaAlaAlaGlyAsnAspProPheLysArgThrThrAlaLeuGluThr 155

Qy    478 TGC------CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTA 531
                ||||||:::||||||   |||      ::::::         |||
Db    156 PheSerLysGlnGluIleAlaLysCysGluGlu---GluGlnGlyTyrArgCysGlnPhe 174

Qy    532 GAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTT 591
                |||      |||   |||   :::                  ||||||||
Db    175 PheSerPheAlaGlyGlyAsnThrTyrArgValPheLysAsnLeuGluIleLysAspVal 194

Qy    592 CGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGG 651
          |||:::|||:::||||||         ||||||||||||||||||||||   |||||||||
Db    195 ArgLeuValTyrAlaProGlnGlySerValGlyLysPheGlyGlyAspValAspAsnTrp 214

Qy    652 ATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAAC 711
          |||||||||||||||||||||||||||   :::|||   |||   |||   |||
Db    215 MetTrpProArgHisThrGlyAspPheSerPheTyrArgAlaTyrValGlyLysAspGly 234

Qy    712 CGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTA 771
          :::||||||   :::||||||:::|||   ||||||:::|||   ::::::
Db    235 LysProAlaSerPheSerLysGluAsnIleProTyrArgProLysHisTrpLeuLysPhe 254

Qy    772 TCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACG 831
          |||   |||               |||:::   |||      |||:::||||||   |||
Db    255 SerAspGlnProLeuGlyAspGlyAspPheValMetValAlaGlyTyrProGlyArgThr 274

Qy    832 GATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGT 891
          :::||||||                                 |||   ||||||
Db    275 AsnArgTyrAlaLeuVal-----------------AlaGluPheGluAsnThrAlaHis 288

Qy    892 ATCGAAGTTCGCGGTATCAAGCAA-----------------GGCATCTGGAAGGAAGCC 933
                   |||   |||                  :::   :::   |||
Db    289 TrpThrTyrProValIleGlyGlnHisPheLysAsnLeuIleAlaLeuIleGluAlaAla 308

Qy    934 ATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAAC 993
                :::            ::::::|||||||||||         |||         |||
Db    309 SerLysGlnAsnProAspIleGlnValLysTyrAlaSerThrLeuAlaGlyLeuAsnAsn 328

Qy    994 TATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGT 1053
             ||||||         |||         |||      |||::::::   ||||||:::
Db    329 ThrSerLysAsnPheAspGlyGlnLeuAspGlyPheArgArgIleAsnAlaIleGlyGln 348

Qy   1054 AAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGC----------- 1101
          |||::::::||||||   |||         |||:::::::::   |||
Db    349 LysGlnSerGluGluThrAlaValLeuAlaTrpLeuLysGlnGlnGlyIleArgGlyHis 368

Qy   1102 AAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCC 1161
          ::::::               |||      |||   :::   |||
Db    369 GluAlaLeuAlaAlaHisGlnThrLeuValAspLeuThrGluGlnTyr------------ 384

Qy   1162 AAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTT 1221
          |||||||||::::::         ::::::         |||   |||:::   |||:::
Db    385 LysAlaAsnGlnAspArgAspPheValLeuGlyGln---PheAsnGlySerGlyValIle 403

Qy   1222 CGTTTTGCACAGTTTGCCAACGCATTGGCT---------ACAAATCCTGATGCTCAT--- 1269
             |||         ||||||         |||   ||||||
Db    404 GlyValAlaValAsnLeuTyrArgLeuAlaIleGluArgThrLysSerAspAlaGlnArg 423

Qy   1270 ---------------------------------GCCGGTATCCTCAAATCGCTTGAC 1293
                                           |||   ||||||   ::::::
Db    424 GluAlaGlyTyrGlnGluArgAspLeuProThrIleGluGlyAsnLeuLysGlnMetGlu 443

Qy   1294 GACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTG----------------- 1335
                :::   |||||||||   :::||||||::::::
Db    444 ---------ArgArgTyrLeuProGluMetAspArgGlnMetGlnGlnTyrTrpLeuThr 460

Qy   1336 -----------CTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCC------ 1377
                     ||||||               |||::::::|||:::   |||
Db    461 GluTyrAsnLysLeuPro--------------ValLysGlnArgValAlaAlaIleAsp 475

Qy   1378 -----------GACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAA 1425
                     |||   :::|||      |||   :::   |||   |||
Db    476 ValTrpLeuGlyAspGlyIleProAlaThrLeuLysArgLeuGlyAspThrLysLeuSer 495

Qy   1426 GGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGAC 1485
             ::::::
Db    496 SerSerGluGluArg--------------------------------------------- 500

Qy   1486 AAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAA 1545
                     ||||||      :::   ::::::   |||            :::
Db    501 ---------------LeuLysTrpPheAsnAlaAspArgAlaAlaPheGluSerSerGln 515

Qy   1546 GATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGAT 1605
          |||||||||:::         :::   :::::::::   |||         |||:::
Db    516 AspProAlaIleArgTyrAlaValAlaIleMetProAlaLeuLeuGluIleGluArgGln 535

Qy   1606 GCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGT 1665
                   :::   |||   :::   ::::::               |||
Db    536 AsnLysIleArgThrGlyGluLeuLeuLysAlaArgProIleTyrLeuGlnAlaLeuAla 555

Qy   1666 GAGATGTACCCC------GGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATG 1719
          :::                  |||:::   :::      |||||||||   ::::::|||:::
Db    556 AspTyrAsnLysSerHisGlyLysPheValTyrProAspAlaAsnSerSerLeuArgIle 575

Qy   1720 AGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACG 1779
          ::::::|||   :::|||||||||   |||:::||||||         |||      |||
Db    576 ThrPheGlyHisValLysGlyTyrSerProLysAspGlyValGluTyrThrProPheThr 595

Qy   1780 ACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAG 1839
          |||   :::|||||||:::   |||   :::         :::   |||         :::
Db    596 ThrLeuGlnGlyValMetAlaLys---AsnThrGlyValGluProPheAspSerProLys 614

Qy   1840 AATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAAC------GGT 1893
          ::::::::::::         :::   |||:::|||         |||:::      |||
Db    615 SerLeuIleAsnAlaIleLysAlaLysSerTyrAlaAsnLeuAlaAspGlnArgIleGly 634

Qy   1894 CAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCC 1953
             :::   :::   |||||||||:::   |||||||||||||||||||||||||||
Db    635 ThrValProValAsnPheLeuSerAspLeuAspIleThrGlyGlyAsnSerGlySerPro 654
```

```
Qy   1954 GTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATG 2013
          |||   |||    :::|||:::|||:::|||||||||||||||||||||||||||::::::
Db    655 ValLeuAspAlaHisGlyLysLeuValGlyLeuAlaPheAspGlyAsnTrpGluSerVal 674

Qy   2014 AGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTAC 2073
          |||   :::       |||:::|||   :::   |||||||||:::||||||    ||||||
Db    675 SerSerAsnTrpValPheAspProValMetThrArgThrIleAlaValAspSerArgTyr 694

Qy   2074 GTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTG 2133
          |||   :::::::::   :::                   |||   |||::::::||||||   |||
Db    695 ValGlnTrpIleMetThrGluValAlaProAlaProHisLeuLeuLysGluLeuAsnLeu 714
```

RESULT 2
T02345
hypothetical protein KIAA0324 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C;Accession: T02345
R;Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.; re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A;Description: Sequencing of human chromosome 16p13.3.
A;Reference number: Z14664
A;Accession: T02345
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1791 <RIC>
A;Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650
C;Genetics:
A;Map position: 16
A;Introns: 1610/2; 1706/2
A;Note: KIAA0324

```
Alignment Scores:
Pred. No.:                 0.0236        Length:          1791
Score:                     156.00        Matches:         184
Percent Similarity:        34.42%        Conservative:    81
Best Local Similarity:     23.90%        Mismatches:      276
Query Match:               4.08%         Indels:          229
DB:                        2             Gaps:            35
```

US-10-008-355-1 (1-2139) x T02345 (1-1791)

```
Qy     27 TCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGGGTAGCCAAAGCCGACAAAGGCATGTG 86
          |||   |||||||||   |||                  |||:::||||||         :::
Db    611 SerGlySerSerProGluVal------------AspSerLysSerArg---------Leu 623

Qy     87 GCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCT 146
          :::|||:::|||::::::||||||   |||                              |||
Db    624 SerProArgArgSerArgSerGlySerSerProGluValLysAspLysProArgAlaAla 643

Qy    147 CCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTT 206
          |||
Db    644 Pro--------------------------------------------------------- 644

Qy    207 CGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCA 266
                                      :::||||||   |||
Db    645 -----------------------ArgAlaGlnSerGlySerAsp-------------- 651

Qy    267 CTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGG 326
                               ||||||||||||         |||         |||
Db    652 -----------SerSerProGluProLysAlaProAlaProArgAlaLeuProArgArg 667

Qy    327 TTTCGT------TTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAA 380
             |||         |||      |||   |||   |||      ||||||
Db    668 SerArgSerGlySerSerSerLysGlyArgGlyProSerProGluGlySerSerSerThr 687

Qy    381 GTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCAC 440
             |||:::   :::                        |||||||||:::
Db    688 GluSerSerProGluHisProProLysSer---------ArgThrAlaArg--------- 701

Qy    441 TGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACT---GGCCAAAAA 497
             |||      |||:::                  ||||||      ::::::

Db    702 ---ArgGlySerArgSerSerProGluProLysThrLysSerArgThrProProArgArg 720

Qy    498 AGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAA------ 551
          |||      |||      |||   |||   |||:::|||   |||   ::::::
Db    721 ArgSerSerArgSerSerProGluLeuThrArgLysAlaArgLeuSerArgArgSerArg 740

Qy    552 -----------------------------CGAATACTTCCT---CATCGTCTACGATGT 578
                                       |||      |||   ||||||
Db    741 SerAlaSerSerSerProGluThrArgSerArgThrProProArgHisArg--------- 757

Qy    579 ATTCAAGGACGTTCGTATGGTATTTGCTCC---------TCCCAGCTCTGTAGGTAAGTT 629
                   ||||||   :::   |||            ::::::      |||
Db    758 ---------ArgSerProSerValSerSerProGluProAlaGluLysSerArgSerSer 774

Qy    630 CGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCG 689
          ||||||||||                  :::||||||   |||   :::
Db    775 ArgArgArgArg--------------SerAlaSerSerProArgThrLysThrThrSer 789

Qy    690 CGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAA 749
          |||      |||            |||      |||
Db    790 ArgArgGlyArgSerProSerProLysProArg-------------------------- 800

Qy    750 GCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGAC 809
                |||   |||   |||      :::|||   :::
Db    801 ---GlyLeuGlnArgSerArgSerArgSerArgArgGluLys------------------ 813

Qy    810 CATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTAT 869
                                                               |||:::
Db    814 ----------------------------------------------ThrArgThrThr 817

Qy    870 CGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGA 929
          |||:::|||:::::::|||              ||||||:::      :::|||      :::
Db    818 ArgArgArgAspArgSerGly------SerSerGlnSerThrSerArgArgArgGlnArg 835

Qy    930 AGCCATGAGCGCAGATCAGGCTACCCGTATCAAA--------------TATGCCAGCAA 974
          |||      |||   :::   ||||||   :::                  |||   |||::
Db    836 SerArgSerArg-SerArgValThrArgArgArgArgGlyGlySerGlyTyrHisSerAr 855

Qy    975 GTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCG 1034
          :                     :::   |||              ||||||   :::||
Db    855 gSerProAlaArgGlnGluSerSerArgThrSerSerArgArgArgArgGlyArgSerAr 875

Qy   1035 TCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTC---GCAGACTGGATCCG 1091
          |               ||||||||||:::         |||            |||   |||   ||
Db    875 gThrProProThrSerArgLysArgSerArgSerArgThrSerProAlaProTrpLysAr 895

Qy   1092 TAAGAACGGCAAGAGTG-----------------CTGTCTATGGCGATGTATTGTCTTC 1133
          |      :::|||                      |||   :::         |||:
Db    895 gSerArg-SerArgAlaSerProAlaThrHisArgArgSerArgSerArgThrProLeuI 915

Qy   1134 TCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGA 1193
          ::|||:::|||      |||         |||      |||               :::|
Db    915 leSerArgArgArgSerArgSerArgThrSerProValSerArgArgArgSerArgSerA 935

Qy   1194 GA----------------------------CGCTCTTCGGTGGTACCGAGGTGGTTCG 1223
          ||                            |||:::|||   |||   |||
Db    935 rgThrSerValThrArgArgArgSerArgSerArgAlaSerProValSerArgArgArgS 955

Qy   1224 TTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAA 1283
                |||                                   ||||||:::
Db    955 erArgSerArgThrPro----------------------------ProValThrArgA 965

Qy   1284 ATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGC 1343
             |||      :::   |||   ||||||      ||||||   |||      ||||
Db    965 rgArgSerArgSerArgThrProThrThrArgArgArgSerArgSerArg---ThrProP 984

Qy   1344 CATGCTCGATATTGTACGCCGGCGTATCCCTGCCG------------------------ 1378
          ||   :::         :::         |||
Db    984 roValThrArgArgArgSerArgSerArgThrProProValThrArgArgArgSerArgS 1004

Qy   1379 --ACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAA 1436
             :::|||||||||   |||         :::|||   |||      ||||||
Db   1004 erArgThrSerProIleThrArgArgArgSerArgSerArgThrSerProValThrArgA 1024
```

```
Qy   1437 GA------AGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTT 1490
          ||       :::||||||       ||||||                          |
Db   1024 rgArgSerArgSerArgThrSerProValThrArg------------------ArgArgS 1038

Qy   1491 CCATGCCATGCTCAAGTCCATGGACAAGG---------------AAAAGTTTGCCAAGGC 1535
          ||        :::||||||  ||||||                 ::::::  |||   :
Db   1038 erArgSerArgThrSerProValThrArgArgArgSerArgSerArgThrProProAlaI 1058

Qy   1536 TATCGAGAAAGATCCGGCAGTAGAGC------TTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
          ::   |||:::   |||      :::           ||||||
Db   1058 leArgArgArgSerArgSerArgThrProLeuLeuProArg------------------ 1071

Qy   1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
                   :::      |||        |||:::   :::||||||    |||
Db   1072 --------LysArgSerArgSerArgSerProLeuAlaIleArgArgArgSerArgSerA 1089

Qy   1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCAC 1709
             |||      |||                         |||::::::     ||||
Db   1089 rgThrProArgThrAlaArgGlyLys---------------ArgSerLeuThrArgSerP 1104

Qy   1710 CATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAA 1769
          ||                  |||:::     :::   ||||||:::      |||:::
Db   1104 ro------------------ProAla------IleArgArgArgSerAlaSerGlySerS 1116

Qy   1770 CTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGC 1829
          ::       |||                                  |||:::   :::   |
Db   1116 erSerAspArgSer------------------------------ArgSerAlaThrProP 1126

Qy   1830 CGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAA 1889
          ||      |||      |||   |||      |||                   |||
Db   1126 roAlaThrArgAsnHisSerGlySerArgThrPro------------------ProValA 1140

Qy   1890 CGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAG 1949
            ::::::|||      :::|||      :::     |||   :::   ||||||:::
Db   1140 laLeuAsnSerSerArgMetSerCysPheSerArgProSerMetSerProThrProLeuA 1160

Qy   1950 CCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGC 2009
                                                     :::   |||    |
Db   1160 sp-----------------------------------ArgCysArgSerProGlyMetL 1168

Qy   2010 TATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGA------ 2063
            ||      :::   ||||||      |||:::         |||  |||    |||
Db   1168 euGluProLeuGlySerSerArgThrProMetSerValLeuGln-GlnAlaGlyGlySer 1187

Qy   2064 ---------CATCCGCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCT 2114
                         |||        :::   |||   |||        |||||||||    :::
Db   1188 MetMetAspGlyProGlyProArgIleProAspHisGlnArgThrSerValProGluAsn 1207

Qy   2115 CATCCAAGAGCTGAAGTTGATCTA 2138
           |||   ::::::   :::   |||
Db   1208 HisAlaGlnSerArgIleAlaLeu 1215
```

RESULT 3
S25345
probable membrane protein YCR089w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YCR1102
C;Species: Saccharomyces cerevisiae
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000
C;Accession: S25345; S19504
R;Wilson, C.; Grisanti, P.; Frontali, L.
Yeast 8, 569-575, 1992
A;Title: The complete sequence of a 6146 bp fragment of Saccharomyces cerevisiae chromos
A;Reference number: S25345; MUID:92397594; PMID:1523889
A;Accession: S25345
A;Molecule type: DNA
A;Residues: 1-1609 <WIL>
A;Cross-references: GB:X59720; EMBL:S43845; NID:g1907116; PIDN:CAA42254.1; PID:g1907227
R;Frontali, L.; Grisanti, P.
submitted to the Protein Sequence Database, March 1992
A;Reference number: S19504
A;Accession: S19504
A;Molecule type: DNA
A;Residues: 1-1609 <FRO>
A;Cross-references: EMBL:X59720; NID:g1907116; PID:e264634; PID:g1907227; MIPS:YCR089
C;Genetics:
A;Gene: SGD:FIG2
A;Cross-references: SGD:S0000685; MIPS:YCR089w
A;Map position: 3R
C;Keywords: transmembrane protein
F;4-20/Domain: transmembrane #status predicted <TM1>
F;1592-1609/Domain: transmembrane #status predicted <TM2>

Alignment Scores:

Pred. No.:	0.0285	Length:	1609
Score:	154.50	Matches:	153
Percent Similarity:	35.18%	Conservative:	101
Best Local Similarity:	21.19%	Mismatches:	254
Query Match:	4.04%	Indels:	214
DB:	2	Gaps:	28

US-10-008-355-1 (1-2139) x S25345 (1-1609)

```
Qy     74 ACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGC 133
          |||:::   |||         |||                          ||||||
Db    788 ThrGluValCysSerHisSer--------------------------GluCysThrPro 798

Qy    134 TCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATG 193
          :::   :::   |||      :::   :::|||   ||||||                :::
Db    799 ThrValIleThrSer------ValThrAlaThrSerSerThr---------------Ile 811

Qy    194 CCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCT 253
          |||   ||||||:::                                          |||
Db    812 ProLeuLeuSerThr------------------------------------------Ser 817

Qy    254 TTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAA------GCACGGTGGATC 307
                |||      :::   ||||||         ||||||      |||
Db    818 SerSerThrValLeuSerSerThrValSerGluGlyAlaLysAsnProAlaAlaSerGlu 837

Qy    308 ACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGG 367
            ||||||      ||||||         :::         :::
Db    838 ValThrIleAsnThrGlnValSerAlaThrSerGluAlaThrSerThr------------ 853

Qy    368 GTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGC 427
                                                                   |||
Db    854 ---------------------------------------------------------Ser 854

Qy    428 TCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAAC 487
          ::::::||||||   |||
Db    855 ThrGlnValSerAlaThr------------------------------------------ 860

Qy    488 TGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCA 547
                             |||   |||   |||   |||   :::   |||
Db    861 ---------------SerAlaThrAlaThrAlaSerGluSerSerThrThrSerGlnVal 875

Qy    548 ACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTC------GTATGGTAT 601
          :::|||   :::   :::   ||||||:::   ::::::   |||
Db    876 SerThrAlaSerGluThrIleSerThrLeuGlyThrGlnAsnPheThrThrThrGlySer 895

Qy    602 TTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGC 661
          ||||||   |||||||||      |||:::|||   |||   ||||||        |||
Db    896 LeuLeuPheProAlaLeu------SerThrGluMetIleAsnThrThrValValSerArg 913

Qy    662 GTCACACGGGCGACTTCA------------GCGTATTCCGCGTGTATGCCGGTGCCGACA 709
             |||         |||                   ::::::|||   |||:::|||
Db    914 LysThrLeuIleIleSerThrGluValCysSerHisSerLysCysValProThrValIle 933

Qy    710 ACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCG 769
          |||         :::      |||      |||   :::      :::|||
Db    934 ThrGluValValThrSerLysGlyThr------ProSerAsnGlyHisSerSerGlnThr 951

Qy    770 TATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTA 829
                :::|||      :::|||                  :::|||:::   :::
Db    952 LeuGlnThrGluAlaValGluValThrLeuSerSerHisGlnThrValThrMetSerThr 971

Qy    830 CGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGA--------- 880
```

```
                :::   ::::::   :::             ::::::::       :::
Db     972 GluValCysSerAsnSerIleCysThrProThrValIleThrSerValGlnMetArgSer 991

Qy     881 ---------ACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAG 931
                    |||  |||    :::  ||||||  ||||||  :::|||
Db     992 ThrProPheProTyrLeuThrSerSerThrSerSerSerLeuAlaSerThrLysLys 1011

Qy     932 CCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTA 991
                                    |||:::::  ||||||
Db    1012 -------------------------SerSerLeuGluAlaSer-------------- 1017

Qy     992 ACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTC 1051

Db    1017 ------------------------------------------------------------ 1017

Qy    1052 GTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTG 1111
                                       |||  ||||||  |||  :::  |||
Db    1018 -----------------SerGluMetSerThrPheSerValSerThr---GlnSerLeu 1030

Qy    1112 TCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACC 1171
              :::|||    |||                                |||  |||  |||
Db    1031 ProLeuAlaPheThrCys-----------------------SerGluLysArgSerThr 1042

Qy    1172 GTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCAC 1231
                                       :::  |||  :::|||
Db    1043 ------------------------------ThrSerValSerGlnTrp----------- 1048

Qy    1232 AGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTG 1291
                    ::::::  :::            :::|||    |||||||||  :::
Db    1049 ---------SerAsnThrValLeuThrAsnThrIleMetSerSerSerSerAsnValIle 1065

Qy    1292 ACGACAAGTACAAAG---------ACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCG 1342
           :::|||:::  |||          |||||||||||  :::
Db    1066 SerThrAsnGluLysProSerSerThrThrSerProTyrAsn----------------- 1079

Qy    1343 CCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGA 1402
                                 :::|||  |||||||||:::||||||
Db    1080 -----------PheSerSerGlyTyrSerLeuProSerSerSerThrProSerGlnTyr 1095

Qy    1403 ATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACA 1462
                  ||||||      :::    |||      ::::::|||  |||:::|||
Db    1096 SerLeuSerThr------AlaThrThrThrIleAsnGlyIleLysThrValTyrThrThr 1113

Qy    1463 AGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAA 1522
                  |||      :::|||    |||      :::|||    :::  :::
Db    1114 ------TrpCysProLeuAlaGluLysSerThrValAlaAlaSerSerGlnSerSerArg 1131

Qy    1523 AGTTTGCCAAGGCTA------TCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAA 1576
           |||:::  |||           |||  |||  |||:::     |||||||::::::   |
Db    1132 SerValAspArgPheValSerSerSerLys--ProSerSerSerLeuSerGlnThrSerI 1151

Qy    1577 TAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGG 1636
           ||                                       |||  :::
Db    1151 le--------------------------------------GlnTyrThrLeuSerThrA 1158

Qy    1637 GCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCC------------------ 1677
                       :::|||||||:::  :::|||
Db    1158 laThrThrThr---IleSerGlyLeuLysThrValTyrThrThrTrpCysProLeuThrS 1177

Qy    1678 -------------GGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCT 1723
                        |||              :::  :::       :::|||::::::
Db    1177 erLysSerThrLeuGlyAlaThrThrGlnThrSerSerThrAlaLysValArgIleThr- 1196

Qy    1724 ACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAG 1783
           :::|||   |||                  ::::::     :::|||
Db    1197 SerAlaSerSerAlaThrSerThrSerIleSerLeuSerThrSerThrGluSerGluSer 1216

Qy    1784 GCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATA 1843
           :::     |||              :::
Db    1217 SerSerGlyTyrLeuSerLysGlyValCysSerGlyThrGluCysThrGlnAspValPro 1236

Qy    1844 TCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATA 1903
           :::    :::||||||         |||:::           |||          :::|||
Db    1237 ThrGlnSerSerSerProAlaSerThrLeuAlaTyrSerProSerValSerThrSerSer 1256

Qy    1904 TCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACT-----------------CCG 1945
           |||   |||:::   |||||||||:::        :::|||                |||
Db    1257 SerSerSerPheSerThrThrThrAlaSerThrLeuThrSerThrHisThrSerValPro 1276

Qy    1946 GTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGG 2005
           :::    |||   |||                  |||:::      |||     :::
Db    1277 LeuLeuProSerSerSer--------------SerIleSerAlaSerSerProSerSer 1291

Qy    2006 AAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACA 2065
                    :::   ||||||     :::|||     ::::::   :::    |||
Db    1292 ThrSerLeuLeuSerThrSerLeuProSerProAlaPheThrSerSerThrLeuProThr 1311

Qy    2066 TCCGCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
           :::              ||||||               |||   |||:::||||||||||||
Db    1312 AlaThrAlaValSerSerSerThrPheIleAlaSerSerLeuProLeuSerSerLysSer 1331
```

RESULT 4
I47141
gastric mucin (clone PGM-2A) - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 03-Nov-2000
C;Accession: I47141; S55315
R;Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Gastroenterology 106, 200, 1994
A;Title: Pig gastric mucin: isolation and characterization of a cDNA clone with a nov
A;Reference number: I47141; MUID:94102478; PMID:7506218
A;Accession: I47141
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-528 <TUR>
A;Cross-references: EMBL:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
R;Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Biochem. J. 308, 89-96, 1995
A;Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.
A;Reference number: S55315; MUID:95275264; PMID:7755593
A;Accession: S55315
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-528 <TU2>
A;Cross-references: GB:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
C;Superfamily: pig submaxillary mucin

```
Alignment Scores:
Pred. No.:                 0.0419      Length:          528
Score:                     149.00      Matches:         154
Percent Similarity:        34.22%      Conservative:    91
Best Local Similarity:     21.51%      Mismatches:      243
Query Match:               3.90%       Indels:          228
DB:                        2           Gaps:            28
```

US-10-008-355-1 (1-2139) x I47141 (1-528)

```
Qy      68 AAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGC 127
           :::|||:::   :::        |||   |||   |||
Db       5 GlnProSerSerSerSerSerSerProThrThrSer----------------------- 16

Qy     128 GTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTG 187
              ::::::::::::|||                        |||:::|||      :::
Db      17 ---ThrThrSerValGlnSer--------------------SerSerSerSerVal 28

Qy     188 CCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCC 247
           |||:::|||        :::||||||                   |||      :::
Db      29 ProIleProSerThrThrSerValGln--------------------ProSerSerSer 41

Qy     248 TGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATC 307
              |||   ||||||||:::|||   :::|||
Db      42 GlySerAlaProThrThrSerAlaThrSerVal-------------------------- 52

Qy     308 ACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGG 367

Db      52 ------------------------------------------------------------ 52
```

```
Qy    368 GTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGC 427
                                               :::|||              |||
Db     53 -------------------------------------GlnThrSerSerSerSerSer 59

Qy    428 TCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAAC 487
                :::|||   |||
Db     60 ProProIleSerSerThr------------------------------------------ 65

Qy    488 TGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCA 547
                         :::      :::|||:::|||:::|||        |||:::|||
Db     66 ---------------IleSerValGlnThrSerSerSerSer---------SerValPro 77

Qy    548 ACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTC 607
          ||||||:::|||:::|||                                       :::
Db     78 ThrThrSerThrThrSer---------------------------------------Val 84

Qy    608 CTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACA 667
             |||:::         ||||||   :::      ||||||         |||   |||
Db     85 GlnProSer---------SerSerSerSerAlaProThrThr---------ArgAlaThr 98

Qy    668 CGGGCGACTTCAGCGTATTCCGCGTGTATGCCG--------------------------- 700
                   |||:::   |||:::     |||
Db     99 SerValGlnSerSerSerSerSerSerAlaProIleSerSerThrThrSerValGlnPro 118

Qy    701 ---------------GTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCT 745
                         ||||||||||||         :::         :::   :::
Db    119 SerSerSerGlySerValProThrThrSerAlaThrSerValGlnSerSerSerSerSer 138

Qy    746 ATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCA 805
             :::|||   ||||||                     :::   :::   ::::::
Db    139 SerAlaProThrThrSer------------AlaThrSerValGlnProSerSerSerSer 154

Qy    806 TGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATC 865
             |||   :::|||      |||   :::      |||
Db    155 SerProProIleSerSerThrValSerValGlnProSer--------------------- 167

Qy    866 GTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGA 925
             |||   :::                 |||      :::|||         :::|||
Db    168 ---SerSerSerSerAlaProThrThrSerAlaThrSerValGlnProSerSerSerSer 186

Qy    926 AGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGA 985
                |||         |||         ||||||         :::|||
Db    187 SerProPro---------IleSerSerThrValSerValGlnThrSerSerSerSerSer 203

Qy    986 GTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGA 1045
          |||   |||                        |||:::   :::   |||
Db    204 ValProThr------------------------ThrSerThrThrSerValGlnProSer 215

Qy   1046 TAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGA 1105
                |||||||||               :::         |||||||||::::::
Db    216 SerSerSerSerValProThr---------ThrSerAlaThrSerValArgSerSerSer 232

Qy   1106 GTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGG 1165
                |||      :::              |||   :::
Db    233 SerSerSerThrProIleProSerThrThrSerValGlnProSerSerSerSerSerAla 252

Qy   1166 CCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTT 1225
          ||||||                           :::||||||   |||
Db    253 ProThr------------------ThrSerAlaThrSerValGlnProSerSerSerSer 266

Qy   1226 TTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAAT 1285
                   :::|||:::                     :::   |||   ||||||:::
Db    267 SerThrProIleProSer------------ThrThrSerValGlnProSerSerSerSer 282

Qy   1286 CGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCA 1345
                   |||   |||      ||||||   :::   :::            ||||||
Db    283 SerAlaProThr---ThrSerAlaThrSerValGlnProSerSerSerSerSerProPro 301

Qy   1346 TGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATG 1405
             |||         :::   :::|||:::      ||||||
Db    302 IleSer---------SerThrIleSerValGlnProSerSer------------------ 312

Qy   1406 TAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGA 1465
             |||:::   :::         :::      |||:::|||   |||      |||
Db    313 ---SerSerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGlySer 331

Qy   1466 GTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGT 1525
                      |||||||||         |||   ||||||                  |||
Db    332 AlaProThrThrSerAlaThrSerValGlnProSerSerSer---------------Ser 346

Qy   1526 TTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTG 1585
             |||   :::|||         |||
Db    347 SerProProIleSerSerThrIle------------------------------------ 354

Qy   1586 CTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTC 1645
                                                    ::::::   :::|||
Db    355 ---------------------------------------SerValGlnProSerSerSer 361

Qy   1646 TTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACT 1705
             |||   |||               |||      |||         :::      :::
Db    362 SerSerSerPro------ThrThrSerThrThrSerValGlnProSerSerSerGlySer 379

Qy   1706 TCACCATGCGTATGAGCTACGGCT------CCATCAAGGGATATGAACCGCAGGACGGTG 1759
          :::|||         |||||||:::            ||||||         :::      |||
Db    380 AlaProThrThrSerAlaThrSerValGlnProSerSerSerSerValProThrThr 399

Qy   1760 CCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCG 1819
                |||::::::                        ||||||                 :::
Db    400 SerAlaThrSerVal-----------------------ArgSerSerSerSerSer 411

Qy   1820 ATGAGTTTGCCGTACAGG---------AGAATATCCTCGACCTCTTCCGCACCAAAAACT 1870
                :::|||               :::   ||||||:::||||||         |||   |||
Db    412 ThrProIleProThrThrThrSerValGlnProSerSerSerSerSerValProThrThr 431

Qy   1871 ATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCA 1930
                         :::|||   ||||||   |||               :::|||||||||
Db    432 SerAlaThrSerValGlnThrSerSerSerSerSerThrProIleProSerThrThrSer 451

Qy   1931 --------------------------------CGGGCGGTAACTCCGGTAGCCCCGTAT 1957
                                              :::   |||   |||   |||
Db    452 ValGlnProSerSerSerSerSerAlaProThrThrSerAlaThrSerValGlnProSer 471

Qy   1958 TCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTG 2017
          |||      :::         |||      :::|||:::
Db    472 SerSerSerSerProProIleSerSerThrIleSerValGlnProSer----------- 487

Qy   2018 GTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTC 2077
             :::|||||||||:::|||         |||      |||            |||::::::
Db    488 ---SerSerSerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGly 506

Qy   2078 TCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
          |||:::         |||:::      |||         ||||||   |||
Db    507 SerAlaProThrThrSerAlaThrSerValGlnProSerSerSerSer 522
```

RESULT 5
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR01
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C;Accession: S48478; A26877; B26877; S27281; JC6123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009;
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600; PMID:3106330
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>

```
A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A;Accession: B26877
A;Molecule type: DNA
A;Residues: 762-1331 <YA2>
A;Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A;Reference number: S27281; MUID:89031230; PMID:3141213
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A;Reference number: JC6123; MUID:96323237; PMID:8710886
A;Accession: JC6123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A;Cross-references: MIPS:YIR019c; SGD:S0001458
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F;5-21/Domain: transmembrane #status predicted <TM1>
F;1350-1366/Domain: transmembrane #status predicted <TM2>

Alignment Scores:
Pred. No.:                0.118       Length:         1367
Score:                    145.00      Matches:        160
Percent Similarity:       33.69%      Conservative:   95
Best Local Similarity:    21.14%      Mismatches:     354
Query Match:              3.80%       Indels:         148
DB:                       1           Gaps:           25

US-10-008-355-1 (1-2139) x S48478 (1-1367)

Qy     65 CCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAA 124
             |||   ||||||  :::     ::::::   :::|||       :::
Db    355 ProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSer 374

Qy    125 TGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCA 184
                   |||:::                     :::|||   |||:::|||   |||
Db    375 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 394

Qy    185 TTGCCAATGCCGTGGTTATCTTCGGTG------GCGGATGTACCGGTATCACAGTGTCCG 238
            :::|||   |||       |||:::          :::       |||||||:::
Db    395 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThr 414

Qy    239 ATCAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCA 298
                :::   |||   |||   |||::::::   |||        |||   |||
Db    415 ThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProVal 434

Qy    299 CGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTC 358
                   ||||||      :::                 :::            |||
Db    435 ThrSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSer 454

Qy    359 CGATTCCGGGTC---TTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGG 415
                   |||      |||   |||          |||
Db    455 SerAlaProValProThrProSerSerSerThrThrGluSer------------------ 468

Qy    416 TAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGG 475
                   :::         |||   |||         |||        |||   :::
Db    469 ---SerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProValProThr 487

Qy    476 TATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGC 535
                :::   :::         :::                  |||:::|||    :::   |||
Db    488 ProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSer 507

Qy    536 CTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTA 595
            |||   |||               |||||||||:::|||          |||
Db    508 SerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAla------ 525

Qy    596 TGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGT 655
                                    ||||||       ||||||               ::::::
Db    526 ---------------ProAlaProThrProSerSerSerThrThrGluSerSerSerAla 540

Qy    656 GGCCGCGTCACACGGGCGACTTCAGCGTATTCGCGTGTATGCCGGTGCCGACAACCGGC 715
                  :::       |||    :::    ||||||         |||||||||||
Db    541 ProValThrSerSerThrThrGluSerSerSerAla------ProValProThrProSer 558

Qy    716 CGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACT------------TCG 763
                  |||       :::       |||::::::       |||              |||
Db    559 SerSerThrThrGluSerSerSerThrProValThrSerSerThrThrGluSerSerSer 578

Qy    764 CTGCCGTATCCA---------TGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCG 814
            |||   |||                   |||         :::::    |||   |||:::
Db    579 AlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProValProThr 598

Qy    815 GTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAA 874
            |||   :::            ::::::|||                         |||
Db    599 ProSerSerSerThrThrGluSerSerSerAlaProAlaProThrProSerSerSerThr 618

Qy    875 ACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCA 934
            |||   :::                  :::   ::::::::::::|||
Db    619 ThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSer--------- 635

Qy    935 TGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACT 994
               |||   :::      |||   |||:::                                     |||
Db    636 ---AlaProValProThrProSerSerSer-----------------------ThrThr 646

Qy    995 ATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTA 1054
                             |||   |||      |||
Db    647 GluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAla 666

Qy   1055 AGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCC------------GTAAGAACGGCA 1102
            ||||||         |||   |||   |||              |||   ::::::
Db    667 ProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSer 686

Qy   1103 AGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCA 1162
                  |||   :::                  |||                   |||
Db    687 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 706

Qy   1163 AGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTC 1222
            ||||||                              |||:::   ||||||
Db    707 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValPro--------- 723

Qy   1223 GTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCA 1282
                  |||       |||                           :::      |||   ||||||
Db    724 ThrProSerSerSerThrThrGluSerSerSerAlaProValProThrProSerSerSer 743

Qy   1283 AATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCG 1342
                  ::::::|||      ||||||         :::              |||
Db    744 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 763

Qy   1343 ---------CCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATA 1393
                  |||   |||                  |||   |||      :::|||
Db    764 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValProThrProSer 783

Qy   1394 TATTCAAGAATGTAATCG-----------------ACAAGAAATTCAAAGGCGACACGA 1435
            |||         |||                  |||   :::|||      |||
Db    784 SerSerThrThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGlu 803

Qy   1436 AGAAGTATG-----------CAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCG 1483
            |||:::                  |||   |||:::   :::
Db    804 SerSerValAlaProValProThrProSerSerSerSerAsnIle-------------- 818

Qy   1484 ACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGA 1543
            |||||||||   |||   |||:::|||::::::      :::
Db    819 ThrSerSerAlaProSerSerThrProPheSerSerSerThr----------------- 832

Qy   1544 AAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCG 1603
```

```
Db     833 -----------------------------------------------------GluSer 834

Qy    1604 ATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGC 1663
                   |||:::|||   |||             :::    |||
Db     835 SerSerValProValProThrProSerSerSerThrThrGluSerSerSerAlaProVal 854

Qy    1664 GTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACT---------TCACCATGC 1714
                   |||    :::              ::::||||||         |||
Db     855 SerSerSerThrThrGluSerSerValAlaProValProThrProSerSerSerSerAsn 874

Qy    1715 GTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCT------------ 1762
           :::   ::::::|||||||||                   ::::::|||
Db     875 IleThrSerSerAlaProSer-------------------SerIleProPheSerSerThr 888

Qy    1763 ------------------GGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGC 1804
                             |||||||||:::            :::      |||      |||
Db     889 ThrGluSerPheSerThrGlyThrThrValThrProSerSerSerLysTyrProGlySer 908

Qy    1805 AGGATCCTAAGAGCGATGAGT--------------------TTGCCGTACAGGAGAATA 1843
           :::          ::::::|||                     :::|||    :::
Db     909 GlnThrGluThrSerValSerSerThrThrGluThrThrIleValProThrLysThrThr 928

Qy    1844 TCCTCGACCTCTTCCGCACCAAAAACTATGGTC-------------------------- 1876
           ::::|||    ::::::              |||    :::
Db     929 ThrSerValThrThrProSerThrThrThrIleThrThrThrValCysSerThrGlyThr 948

Qy    1877 --------------------------GCTATGCCGAGAACGGTCAGCTCCATATCGCTT 1909
                                     |||:::||||||||:::::
Db     949 AsnSerAlaGlyGluThrThrSerGlyCysSerProLysThrValThrThr--------- 965

Qy    1910 TCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACG 1969
           :::        |||||||||:::   :::||||||        :::      :::|||   |||
Db     966 ThrValProThrThrThrThrThrSerValThrThrSerSerThrThrThrIleThrThr 985

Qy    1970 GCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGT 2029
              |||                |||      |||                   ||||||
Db     986 ThrVal-----------CysSerThrGlyThrAsnSerAlaGlyGluThrThrSerGly 1001

Qy    2030 TCGAACCCG--------------------ATCTGCAGCGCACAATCAGCGTGGACATCC 2068
              :::|||                     ||||||                 |||
Db    1002 CysSerProLysThrIleThrThrThrValProCysSerThrSer---------ProSer 1018

Qy    2069 GCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCC 2119
              |||   |||          |||         :::    ||||||:::::
Db    1019 GluThrAlaSerGluSerThrThrThrSerProThrThrProValThrThr 1035
```

RESULT 6
S44784
C30C11.4 protein - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 29-Sep-1999
C;Accession: S44784
R;Favello, A.D.
submitted to the EMBL Data Library, January 1993
A;Description: Sequence of the C. elegans cosmid C30C11.
A;Reference number: S44782
A;Accession: S44784
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-776 <FAV>
A;Cross-references: EMBL:L09634; NID:g156220; PIDN:AAA27967.1; PID:g156223
C;Genetics:
A;Introns: 36/2; 145/3; 255/1; 417/2; 732/3
C;Superfamily: heat shock protein 91

Alignment Scores:

Pred. No.:	0.117	Length:	776
Score:	143.50	Matches:	118
Percent Similarity:	37.48%	Conservative:	87
Best Local Similarity:	21.57%	Mismatches:	189
Query Match:	3.76%	Indels:	153
DB:	2	Gaps:	27

US-10-008-355-1 (1-2139) x S44784 (1-776)

```
Qy     113 ATCTGGATCGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCG 172
            ||||||   :::                ||||||             |||
Db     236 IleTrpPheAsp-----------AlaLeuIleArgGluHisPheArgLysGluPheLys 251

Qy     173 ACAAGCCGTCCATTG--------------CCAATGCCGTGGTTATCTTCGGTGGCGGAT 217
            |||        :::                |||  |||||||||      :::  :::
Db     252 ThrLysTyrGlyIleAspAlaAlaThrSerProArgProTrpLeuArgLeuLeuAspGlu 271

Qy     218 GTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCACTGCGGATACG 277
                         :::     :::   |||        |||    ||||||   |||
Db     272 CysGluArgValLysLysGlnMetSerAlaAsnGlnThrProIle-ProLeuAsnIleGl 291

Qy     278 GTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTC 337
            |||:::   |||   |||               ::::::   |||      |||
Db     291 uCysPheMetGluAspLysAspValThrGlyLysMetGln-ArgGlnGluPhe------- 308

Qy     338 GCACGATGGGTGAGGAGCTTCCGATTCCG------------------------------ 366
                          |||:::|||      |||
Db     309 ----------GluAspLeuAlaAlaProIlePheAsnArgIleLysGlnValLeuIleA 325

Qy     367 -------------GGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACA 412
                          |||:::|||:::|||                 :::      |||:
Db     325 snLeuPheAlaAspGlyValSerIleLys--------------ProGluGluIleAspG 340

Qy     413 AGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGG 472
            ::::::|||                |||                |||:::      :::|
Db     340 luIleGlu---------------IleValGlyGlySerSerArgIleProMetIleArgG 355

Qy     473 AGGTATGCCAAGAACTGGCCAAAAAAAGAA--------------AATGCAGACGAG---- 513
            ||:::   ::::::|||          ||||||             |||   ||||||
Db     355 luIleValLysAspLeuPheGlyLysGluProLysThrThrMetAsnGlnAspGluAlaV 375

Qy     514 ----------------AACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAAT 556
                            |||       |||:::   |||      :::         |||:
Db     375 alAlaArgGlyAlaAlaMetGlnCysAlaIleLeuSerProThrPheArgValArgGluP 395

Qy     557 ACTTCCTCATCGTCTACGATGTATTCAAGGAC--------------GTTCGTATGGTAT 601
            ::                         ||||||              :::|||:::   :
Db     395 he-----------------AlaIleLysAspThrGlnProTyrArgIleArgLeuSerT 409

Qy     602 TTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGC 661
            ::            :::|||   |||:::   ||||||:::   |||     :::     ||||
Db     409 rp---------AsnSerThrGlyGluAsnGlyGlyGluAsnAspValPheSer---ProA 425

Qy     662 GTCACACGGGCGACTTCAGC------GTATTCCGCGTGTATGCCGGTGCCGACAACCGGC 715
            ||            ||||||      |||         :::   ::::|||      |||
Db     425 rgAspGluValProPheSerLysLeuValSerLeuLeuArgSerGlyProPheAsnValG 445

Qy     716 CGGCCGAATACAGCAAGGACAAT---AAACCCTATAAGCCCGTTTACTTCGCTGCCGTAT 772
              |||   |||::::::      |||         |||:::      |||:::       :::
Db     445 luAlaHisTyrAlaGlnProAsnValValProHisAsnGlnValHisIleGlySerTrpL 465

Qy     773 CCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGG 832
              :::   |||    :::                               ||||||:::
Db     465 ysValAsnGlyAlaArg----------------------------ProGlyAlaAsp- 474

Qy     833 ATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTA 892
                                      |||         ::::::                 ::::
Db     475 -------------------GlyGlyAsnGlnLysVal-----------------LysV 482

Qy     893 TCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTA 952
            ::::::||||||            ||||||:::      |||
Db     482 alLysValArgValAsnProAspGlyIlePheThrIleAla------------------ 495

Qy     953 CCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCG 1012
                                              ||||||      ::::::     |||
Db     496 -------------------------------SerAlaThrMetTyrGluProArgIleV 505

Qy    1013 GTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAG 1072
                                      ::::::|||   |||      :::   |||      |
```

```
Db     505 alGluGluValProAlaGluAlaMetGluValAspGlyAspAlaLysThrGluAlaProA 525

Qy    1073 CATTC---GCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGT 1129
           ||        :::   :::::|||                  |||      :::|||:::
Db     525 laGluProLeuGluProValLysLysThrLysLeuValProValAspLeuGluValIleG 545

Qy    1130 CTTCTCTCGAAAAGGCTTAT---------------------------------------- 1149
              |||:::       :::|||
Db     545 luSerIleProValSerTyrAspValGlnLysPheHisAsnLeuGluLeuGlnMetGlnG 565

Qy    1150 ----------AAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGC 1198
                     :::|||   |||||||||:::   :::   :::   |||   |||    :
Db     565 luSerAspAlaArgGluLysAlaLysAlaAspAlaLysAsnSer---LeuGluGluTyrV 584

Qy    1199 TCTTCGGTGGTACCGAGGTGGTT------CGTTTTGCACAGTTTGCCAACGCATTGGCTA 1252
           :::::        :::   |||        ::::::|||:::|||            |||
Db     584 alTyrGluMetArgAspLysValSerAspGlnTyrAlaGluPheIleThrProAlaAlaA 604

Qy    1253 CAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACC 1312
                  |||            :::|||   |||   :::|||            |||
Db     604 la------AspGluPheArgSerValLeuThrSerThrGluAspTrpLeuTyrAspGluG 622

Qy    1313 TCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCC 1372
                     :::|||                          |||:::      :::|||:::
Db     622 lyGluAspAlaGluArg--------------------AspValTyrGluLysArgLeu- 634

Qy    1373 CTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACA 1432
                     :::|||         :::           |||:::::::::    :::    :::|
Db     635 -----SerGluLeuLysAlaValGlyThrProValValGluArgTyrArgGluSerGluT 653

Qy    1433 CGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTT-----------CCTTATA 1480
            |||:::|||   |||          ||||||:::|||::::::             |||
Db     653 hrArgLysProAlaPheAspSerPheAspGlnSerIleMetArgValArgLysAlaTyrG 673

Qy    1481 GCGAC--------------AAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGT 1525
             |||                     :::|||   |||   |||   :::   ||||||
Db     673 luAspTyrAlaAsnGlyGlyProThrTyrAlaHisLeuAspSerLysGluMetGluLysV 693

Qy    1526 TTGCCAAGGCTATCGAG 1542
                    |||||||||
Db     693 alIleAsnAlaIleGlu 698
```

RESULT 7
T11678
hypothetical protein SPBC21D10.06c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: T11678
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z17313
A;Accession: T11678
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-948 <SEE>
A;Cross-references: EMBL:AL031536; NID:e1319499; PID:e1319505
A;Experimental source: strain 972h(-)
C;Genetics:
A;Map position: IIR
A;Note: SPBC21D10.06c

Alignment Scores:

Pred. No.:	0.19	Length:	948
Score:	141.00	Matches:	134
Percent Similarity:	34.87%	Conservative:	78
Best Local Similarity:	22.04%	Mismatches:	252
Query Match:	3.69%	Indels:	144
DB:	2	Gaps:	24

US-10-008-355-1 (1-2139) x T11678 (1-948)

```
Qy     503 ATGCAGACGAGAACCAACTCTGCATCG---TAGAGCCTTTCTATTCCAACAACGAATACT 559
           :::   :::   |||   |||||||||      :::::::::|||      |||:::
Db      54 LeuAspSerAsnThrThrSerAlaSerValGlnThrIleAlaIleSerGlnThrAspAsn 73

Qy     560 TCCTCA-----------TCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTC 607
           ::::::                 |||:::::::::   |||   :::   |||   |||
Db      74 AlaAlaSerCysIleProSerAlaSerLeuLeuSerSerValValLeuTyrSerAla 93

Qy     608 CTCCCAGCTCTGTAGGTAAGTTCG------GAGGCGATACGGACAACTGGATGTGGCCGC 661
                 :::   |||||||||               :::   |||:::      |||
Db      94 LysGluThrValThrValSerSerTyrTrpSerLeuValSerThrSerValThrGlyThr 113

Qy     662 GTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCG 721
           |||         ::::::   ||||||   |||      |||:::         |||
Db     114 ValTyrValProTyrThrSerSerValAlaCysPheProTyrAlaThrSerAspAlaPro 133

Qy     722 AAT---------------ACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTG 766
           |||                 :::   |||         |||:::|||
Db     134 AsnProIleProArgGlyAspSerAlaThrSerThrSerIleAlaPro----------- 149

Qy     767 CCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCA 826
              |||            :::   ||||||:::            |||   |||
Db     150 ThrTyrSerAlaSerAspSerSerAlaThrThrIleThrSerSerSerProSer------ 167

Qy     827 GTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATC 886
                       ||||||:::::|||                  :::         :::
Db     168 -----------ThrSerIleIleGlyThrGlySerThrAspThrSerValSerSerThr 183

Qy     887 CTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATC 946
           |||         ::::::|||         :::                  :::
Db     184 LeuThrTyrHisThrProIleAlaSerProThrThrSerSerAsn------SerAspAsn 201

Qy     947 AGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATT 1006
                  |||      :::   :::|||   |||               |||      :::
Db     202 GluTyrThrValAspValIleThrSerSerSerLeuSerSerPheValIleThrAsnVal 221

Qy    1007 CGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGG 1066
              |||
Db     222 AspSerThr--------------------------------------------------- 224

Qy    1067 AAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTAT 1126
                                          |||      |||::::::
Db     225 -----------------------------ThrThrSerValIleAsnTyrIleGlyAla 234

Qy    1127 TGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATT 1186
              |||                  :::               ||||||
Db     235 SerThrLeuGluSerSerSerLeuThrAsnThrValSerProThrGluSerThrPheTyr 254

Qy    1187 TGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCAT 1246
              :::   |||:::                              |||:::||||||
Db     255 GluThrLysSerSerThr-----------------------SerSerValProThrGln 266

Qy    1247 TGGCTA---------CAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACA 1297
              :::                                 |||||||||         :::|||
Db     267 ThrIleAspSerSerSerPheThrSerSerThrProValSerLeuThrSerSerSerThr 286

Qy    1298 AGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTG 1357
           |||:::      :::|||      ||||||            |||   :::|||
Db     287 SerSerSerGlySerSerGlnAspSerThrThrIleAspSerThrProSerThrIleAla 306

Qy    1358 TACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATA------------------------ 1393
            :::   :::            |||||||||||||||
Db     307 ThrSerThrLeuGlnProThrThrSerSerProIleThrThrSerAlaProSerLeuSer 326

Qy    1394 -----------------TATTCAAGAATGTAATCGACAAGA------------------ 1417
                            |||         ||||||
Db     327 SerAlaLeuProThrThrTyrProSerSerLeuSerThrGluValGluValGluTyrPhe 346

Qy    1418 AATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTT 1477
                    |||      |||         :::|||||||||   |||         |||
Db     347 ThrLysThrIleThrAspThrSerSerIleValThrTyrSerThrGlyValGluThrLeu 366

Qy    1478 ATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTA 1537
              |||   ::::::   |||                           :::
Db     367 TyrGluThrGluThrIleThrSerSerGlu------IleSerSerIleIleTyrAsnPhe 384
```

```
Qy   1538 TCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTC 1597
          |||      |||      |||||||||                           |||
Db    385 SerThrProIleSerGlySerSerPheProAspGly--------------------Phe 397

Qy   1598 AGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCG 1657
           :::|||:::      |||        |||         :::|||         :::|||
Db    398 LysProIleAsn---ProThrSerPheProSerLeuThrSerSerThrLysLysIlePro 416

Qy   1658 GTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACT---TCACCATGC 1714
                                                 :::||||||   |||
Db    417 -----------------------------SerThrThrLeuProThrSerSerLysMet 426

Qy   1715 GTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATC 1774
           :::     |||   ||||||      ::::::     |||              |||
Db    427 IleThrThrThrThrProSer------ValSerAsnAsnThrGlnSerSerPheLeuIle 444

Qy   1775 ATACGACAG---GCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCG 1831
          |||              :::|||      |||        :::      |||:::|||
Db    445 IleSerThrPheThrSerSerTyrGluHisSerGluProPheLysValSerSerValPro 464

Qy   1832 --------------TACAGGAGAATA------TCCTCGACCTCTTCCGCACCAAAAACT 1870
                         :::      |||       ||||||   ||||||   |||   |||
Db    465 LeuThrSerAsnAsnPheSerSerIleSerHisSerSerAlaSerSerLeuProIleThr 484

Qy   1870 ------------------------------------------------------------ 1870

Db    485 ProSerSerTyrLeuSerAsnThrThrLeuHisSerSerValGlnSerSerGlnSerSer 504

Qy   1871 ------ATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACG 1924
                   |||::::::|||    :::          |||   |||   |||         |||
Db    505 GlnPheThrValSerValProSerSerThrGlnSerTyrSerThrSerSerAsnPheThr 524

Qy   1925 ACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTC 1984
          |||                       |||   :::|||   |||::::::   |||
Db    525 Thr-----------------------ProIleThrIleSerThrSerLeuSerSerPhe 536

Qy   1985 TTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAAC---CCGATC 2041
                ::::::   :::                          :::   |||||||||
Db    537 ProThrThrIleValSerSerSerPheGlnTyrSerSerLeuSerSerAsnValThrThr 556

Qy   2042 TGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGTC 2101
             :::|||||||||:::         |||::::::   ||||||   ||||||:::
Db    557 ThrAsnAlaGlnSerSerSerLeuSerSerSerAsnSerSerAlaLeuThrHisIleSer 576

Qy   2102 AGTGCCCCCGTCTCATCCAAGAGC 2125
          |||:::   |||||||||   |||
Db    577 SerSerIleValSerSerGlySer 584
```

RESULT 8
S69703
HKR1 protein precursor - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YDR420w
C;Species: Saccharomyces cerevisiae
C;Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 23-Mar-2001
C;Accession: S69703; A53382
R;Dietrich, F.S.
submitted to the EMBL Data Library, August 1995
A;Description: The sequence of S. cerevisiae lambda 3641 and cosmids 9461, 9831, and 941
A;Reference number: S69555
A;Accession: S69703
A;Molecule type: DNA
A;Residues: 1-1802 <DIE>
A;Cross-references: EMBL:U33007; NID:g927685; PIDN:AAB64857.1; PID:g927691; MIPS:YDR420w
R;Kasahara, S.; Yamada, H.; Mio, T.; Shiratori, Y.; Miyamoto, C.; Yabe, T.; Nakajima, T.
J. Bacteriol. 176, 1488-1499, 1994
A;Title: Cloning of the Saccharomyces cerevisiae gene whose overexpression overcomes the
A;Reference number: A53382; MUID:94156857; PMID:8113191
A;Accession: A53382
A;Molecule type: DNA
A;Residues: 1-581,'A',583-593,'A',595-1802 <KAS>
A;Cross-references: EMBL:S69101; NID:g545659; PIDN:AAB30051.1; PID:g545660
A;Experimental source: YNN295
A;Note: sequence extracted from NCBI backbone (NCBIN:144410, NCBIP:144411)
C;Genetics:
A;Gene: SGD:HKR1
A;Cross-references: SGD:S0002828; MIPS:YDR420w
A;Map position: 4R
C;Keywords: calcium binding; glycoprotein; transmembrane protein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1802/Product: HKR1 protein #status predicted <MAT>
F;1483-1508/Domain: transmembrane #status predicted <TMM>
F;1645-1656/Domain: calcium binding #status predicted <CA1>

```
Alignment Scores:
Pred. No.:               0.317        Length:         1802
Score:                   139.50       Matches:        168
Percent Similarity:      35.54%       Conservative:   106
Best Local Similarity:   21.79%       Mismatches:     292
Query Match:             3.65%        Indels:         206
DB:                      2            Gaps:           32
```

US-10-008-355-1 (1-2139) x S69703 (1-1802)

```
Qy      1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
            :::   :::|||:::|||   |||||||||         :::
Db      2 ValSerLeuLysIleLysLysIleLeuLeuLeuValSer--------------------- 14

Qy     61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
                                    |||||||||   :::         :::   |||
Db     15 --------------------------LeuLeuAsnAlaIleGluAlaTyrSerAsnAsp 25

Qy    121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
             :::            :::      ::::::|||      |||:::
Db     26 ThrIleTyrSerThrSerTyrAsnAsnGlyIleGluSerThrProSerTyrSerThrSer 45

Qy    181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
          :::|||::::::                          ||||||   :::         :::
Db     46 AlaIleSerSer--------------------------ThrGlySerSerAsnLysGlu 56

Qy    241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGC--- 297
                                                ||||||   |||   |||
Db     57 ------------------------------------AsnAlaIleThrSerSerSerGlu 64

Qy    298 ------ACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAG 351
                |||:::         |||         |||      |||   |||:::   :::
Db     65 ThrThrThrMetAlaGlyGlnTyrGlyGluSerGly------SerThrThrIleMetAsp 82

Qy    352 GAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGAC 411
          |||         |||   |||   :::|||               :::   ||||||
Db     83 GluGlnGlu---ThrGlyThrSerSerGlnTyr------------IleSerValThrThr 97

Qy    412 AAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAG 471
                :::                 :::|||   |||         ::::::|||:::
Db     98 ThrThrGln--------------ThrSerAspThrMetSerSerValLysLysSerThr 112

Qy    472 GAGGTATGCCAAGAACTGGCCAAA-----------------AAAGAAAATGCAGACGAG 513
           |||:::            :::                      :::      :::||||||
Db    113 GluIleAlaThrProSerSerSerIleValProThrProLeuGlnSerTyrSerAspGlu 132

Qy    514 AACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTAC 573
           :::|||      :::          |||:::|||
Db    133 SerGln---------IleSerGlnThrLeuSerHisAsn-------------------- 142

Qy    574 GATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGA 633
                                              |||   ||||||   :::
Db    143 -----------------------------------ProLysSerValAlaGluSerAsp 150

Qy    634 GGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTG 693
             ||||||  |||                         :::|||      |||:::
Db    151 SerAspThr-Thr-------------------------SerSerGluSerSerSerSe 161

Qy    694 TATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCC 753
            :::   :::   |||:::                    |||   |||:::||||||
Db    161 rValIleIleSerThrSerAspSerSerAlaValProArgGluIleSerProIleIleTh 181
```

```
Qy    754 GTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATC 813
                |||              :::    |||        |||            ||
Db    181 rThrAspSerGlnIleSerLysGluGluGlyThrLeuAlaGlnThr---------SerSe 198

Qy    814 GGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAA 873
          |:::|||       ||||||||       :::        |||    :::|||   |||
Db    198 rIleSerGluThrThrArgIleAlaGlnMetValThrArgValSerGlnIleSerSerIl 218

Qy    874 AACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATC----TGGAAGGA 929
           |||              :::     |||:::     |||             :::
Db    218 eThrAlaAlaSerThrIleAspGlyPheSerSerGluSerThrGlnThrAspPheSerAs 238

Qy    930 AGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGC 989
             :::|||   :::    :::        :::||||||        :::|||     ::
Db    238 nThrValSerPheGluAsnSerValGluGluGluTyrAlaMetSerLysSerGlnLeuSe 258

Qy    990 TAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGG 1049
          :      :::   :::|||       :::        |||                  ||
Db    258 rGluSerTyrSerSerSerSerThrValTyrSerGly--------------------Gl 271

Qy   1050 TCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGC 1109
          |        |||:::                          |||         |||
Db    271 yGluSerThrAlaAsp--------------------------LysThrSerSerSerPr 282

Qy   1110 TGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAA 1169
           :::                   |||      :::          |||   ::::::::::
Db    282 oIleThrSerPheSerSerSerTyrSerGlnThrThrSerThrGluThrSerGluSerSe 302

Qy   1170 CCGTGAGATGACTTATTTGAGC------GAGACGCTCTTCGGTGGTACCGAGGTGGTTCG 1223
          :|||           :::|||         ::::::          |||    :::
Db    302 rArgValAlaValGlyValSerArgProSerSerIleThrGlnThrThrSerIleAspSe 322

Qy   1224 TTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGT-------- 1275
            |||:::                 |||:::|||      |||      |||||
Db    322 rPheSerMetSerGluValGluLeuSerThrTyrTyrAspLeuSerAlaGlyAsnTyrPr 342

Qy   1276 ---------------ATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTC 1319
                          :::         |||:::||||||          ||||||
Db    342 oAspGlnGluLeuIleVal-AspArgProAlaThrSerSerThrAlaGluThrSerSerG 362

Qy   1320 GCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGA 1379
            ::::::                                  |||||||||         :
Db    362 luAlaSer--------------------------------GlnGlyValSerArgGluS 371

Qy   1380 CAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAAT---------TCAAAGGCGA 1430
          ::::::::          |||    :::    ||||||    |||         |||    |||:
Db    371 erAsnThrPheAlaValSerSerIle---SerThrThrAsnPheIleValSerSerAlaS 390

Qy   1431 CACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTG--------------------T 1469
          ::          :::    ||||||    :::          |||
Db    390 erAspThrValValSerThrSerSerThrAsnThrValProTyrSerSerValHisSerT 410

Qy   1470 GGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTG- 1528
            |||:::    |||||||||                  ||||||     :::            ||||||
Db    410 hrPheValHisAlaThrSer------------SerSerThrTyrIleSerSerSerLeuT 426

Qy   1529 --------CCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGC 1580
                    |||   ||||||           :::             |||       |||
Db    426 yrSerSerProSerLeuSerAlaSerValSerSer---HisPheGlyValAla------- 442

Qy   1581 TGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAA 1640
                                                     |||    |||            |
Db    443 -----------------------------------ProPheProSerAlaTyrIleS 450

Qy   1641 GCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGC 1700
           ||        |||:::||||||     |||        |||
Db    450 erPheSerSerValProValAlaValSerSerThr--------------------TyrT 463

Qy   1701 CAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGC 1760
            :::||||||                   ||||||         :::          :::||||
Db    463 hrSerSerProSerAlaSerValValValProSerAlaTyrAlaSerSerProSerValP 483

Qy   1761 CTGGTACA---------------ACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCA 1805
                 ||                      |||              |||     |||           |||
Db    483 roValAlaValSerSerThrTyrThrSerSerProSerAlaProAlaAlaIleSerSerT 503

Qy   1806 GGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAA 1865
                 :::    |||    |||                 |||:::    :::|||:::|||
Db    503 hrTyrThrSerSerProSerAlaProValAlaValSerSerThrTyrThrSerSerProS 523

Qy   1866 AAACTATGGTCGCTATGCCGAGAACGGTCAGCTCC--------------------ATAT 1904
                 |||:::       |||    :::|||                             :::|
Db    523 erAlaProAlaAlaIleSerSerThrTyrThrSerSerProSerAlaProValAlaValS 543

Qy   1905 CGCTTTCCTAT-------------------------------CGAACAACGACATCAC 1931
            ||   :::|||                                       |||   ||||||
Db    543 erSerThrTyrThrSerSerProSerAlaProAlaAlaIleSerSerThrTyrThrSerS 563

Qy   1932 GGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTT 1991
                     |||||||||       |||       |||:::       |||         ::::
Db    563 erProSerAlaProValAlaValSerSerThrTyrThrSerSerProSerAlaProValA 583

Qy   1992 CGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCAC 2051
            ::::::::::                  |||||||||    :::             :::
Db    583 laIleSerSer-----------ThrTyrThrSerSerProSerValProValAlaValS 599

Qy   2052 AATCAGCGTGGACATCC--------------GCTACGTTCTCTTCATGATTGACAAATG 2096
               |||    :::||||||                     |||         ||||||       |||:::
Db    599 erSerThrTyrThrSerSerProSerAlaProAlaAlaIleSerSerThrTyrThrSerS 619

Qy   2097 GGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
                 ||||||||||||:::       |||
Db    619 erProSerAlaProValAlaValSerSer 628
```

```
RESULT 9
A53102
alpha-2-macroglobulin receptor precursor - chicken
N;Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein rece
C;Species: Gallus gallus (chicken)
C;Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 22-Jun-1999
C;Accession: A53102
R;Nimpf, J.; Stifani, S.; Bilous, P.T.; Schneider, W.J.
J. Biol. Chem. 269, 212-219, 1994
A;Title: The somatic cell-specific low density lipoprotein receptor-related protein o
A;Reference number: A53102; MUID:94103212; PMID:7506255
A;Accession: A53102
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-4543 <NIM>
A;Cross-references: GB:X74904; NID:g438006; PIDN:CAA52870.1; PID:g438007
C;Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-assoc
d protein.
C;Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-bind
C;Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; glyco
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-3942/Domain: alpha-2-macroglobulin receptor 515K chain #status predicted <515K>
F;18-3942,3943-4543/Product: alpha-2-macroglobulin receptor #status predicted <MAT>
F;29-66/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;74-110/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;117-150/Domain: EGF homology <EG1>
F;156-190/Domain: EGF homology <EG2>
F;200-241/Domain: LDL receptor YWTD-containing repeat homology <YW01>
F;242-283/Domain: LDL receptor YWTD-containing repeat homology <YW02>
F;294-336/Domain: LDL receptor YWTD-containing repeat homology <YW03>
F;337-380/Domain: LDL receptor YWTD-containing repeat homology <YW04>
F;381-422/Domain: LDL receptor YWTD-containing repeat homology <YW05>
F;423-470/Domain: LDL receptor YWTD-containing repeat homology <YW06>
F;480-521/Domain: EGF homology <EG3>
F;573-615/Domain: LDL receptor YWTD-containing repeat homology <YW07>
F;616-661/Domain: LDL receptor YWTD-containing repeat homology <YW08>
F;662-712/Domain: LDL receptor YWTD-containing repeat homology <YW09>
F;713-754/Domain: LDL receptor YWTD-containing repeat homology <YW10>
F;755-797/Domain: LDL receptor YWTD-containing repeat homology <YW11>
F;805-840/Domain: EGF homology <EG4>
F;852-888/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;893-929/Domain: LDL receptor ligand-binding repeat homology <LDL4>
```

F;934-969/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F;974-1009/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F;1013-1049/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F;1060-1095/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F;1102-1138/Domain: LDL receptor ligand-binding repeat homology <LDL9>
F;1143-1180/Domain: LDL receptor ligand-binding repeat homology <LDLA>
F;1183-1219/Domain: EGF homology <EG5>
F;1225-1259/Domain: EGF homology <EG6>
F;1267-1306/Domain: LDL receptor YWTD-containing repeat homology <YW12>
F;1307-1353/Domain: LDL receptor YWTD-containing repeat homology <YW13>
F;1354-1396/Domain: LDL receptor YWTD-containing repeat homology <YW14>
F;1397-1443/Domain: LDL receptor YWTD-containing repeat homology <YW15>
F;1444-1486/Domain: LDL receptor YWTD-containing repeat homology <YW16>
F;1487-1529/Domain: LDL receptor YWTD-containing repeat homology <YW17>
F;1538-1576/Domain: EGF homology <EG7>
F;1581-1624/Domain: LDL receptor YWTD-containing repeat homology <YW18>
F;1625-1667/Domain: LDL receptor YWTD-containing repeat homology <YW19>
F;1668-1711/Domain: LDL receptor YWTD-containing repeat homology <YW20>
F;1712-1751/Domain: LDL receptor YWTD-containing repeat homology <YW21>
F;1752-1794/Domain: LDL receptor YWTD-containing repeat homology <YW22>
F;1795-1842/Domain: LDL receptor YWTD-containing repeat homology <YW23>
F;1846-1882/Domain: EGF homology <EG8>
F;1930-1972/Domain: LDL receptor YWTD-containing repeat homology <YW24>
F;1973-2015/Domain: LDL receptor YWTD-containing repeat homology <YW25>
F;2016-2059/Domain: LDL receptor YWTD-containing repeat homology <YW26>
F;2060-2101/Domain: LDL receptor YWTD-containing repeat homology <YW27>
F;2102-2147/Domain: LDL receptor YWTD-containing repeat homology <YW28>
F;2155-2190/Domain: EGF homology <EG9>
F;2195-2237/Domain: LDL receptor YWTD-containing repeat homology <YW29>
F;2247-2288/Domain: LDL receptor YWTD-containing repeat homology <YW30>
F;2338-2382/Domain: LDL receptor YWTD-containing repeat homology <YW31>
F;2383-2423/Domain: LDL receptor YWTD-containing repeat homology <YW32>
F;2424-2467/Domain: LDL receptor YWTD-containing repeat homology <YW33>
F;2476-2511/Domain: EGF homology <EG10>
F;2518-2555/Domain: LDL receptor ligand-binding repeat homology <LDLB>
F;2560-2594/Domain: LDL receptor ligand-binding repeat homology <LDLC>
F;2599-2633/Domain: LDL receptor ligand-binding repeat homology <LDLD>
F;2646-2682/Domain: LDL receptor ligand-binding repeat homology <LDLE>
F;2690-2724/Domain: LDL receptor ligand-binding repeat homology <LDLF>
F;2732-2767/Domain: LDL receptor ligand-binding repeat homology <LDLG>
F;2772-2810/Domain: LDL receptor ligand-binding repeat homology <LDLH>
F;2816-2851/Domain: LDL receptor ligand-binding repeat homology <LDLI>
F;2856-2895/Domain: LDL receptor ligand-binding repeat homology <LDLJ>
F;2902-2936/Domain: LDL receptor ligand-binding repeat homology <LDLK>
F;2941-2977/Domain: EGF homology <EG11>
F;2983-3018/Domain: EGF homology <EG12>
F;3026-3065/Domain: LDL receptor YWTD-containing repeat homology <YW34>
F;3066-3110/Domain: LDL receptor YWTD-containing repeat homology <YW35>
F;3111-3153/Domain: LDL receptor YWTD-containing repeat homology <YW36>
F;3154-3197/Domain: LDL receptor YWTD-containing repeat homology <YW37>
F;3198-3238/Domain: LDL receptor YWTD-containing repeat homology <YW38>
F;3239-3281/Domain: LDL receptor YWTD-containing repeat homology <YW39>
F;3291-3327/Domain: EGF homology <EG13>
F;3331-3366/Domain: LDL receptor ligand-binding repeat homology <LDLL>
F;3371-3405/Domain: LDL receptor ligand-binding repeat homology <LDLM>
F;3410-3445/Domain: LDL receptor ligand-binding repeat homology <LDLN>
F;3450-3486/Domain: LDL receptor ligand-binding repeat homology <LDLO>
F;3491-3528/Domain: LDL receptor ligand-binding repeat homology <LDLP>
F;3533-3567/Domain: LDL receptor ligand-binding repeat homology <LDLQ>
F;3572-3606/Domain: LDL receptor ligand-binding repeat homology <LDLR>
F;3610-3644/Domain: LDL receptor ligand-binding repeat homology <LDLS>
F;3651-3687/Domain: LDL receptor ligand-binding repeat homology <LDLT>
F;3692-3728/Domain: LDL receptor ligand-binding repeat homology <LDLU>
F;3738-3774/Domain: LDL receptor ligand-binding repeat homology <LDLV>
F;3783-3820/Domain: EGF homology <EG14>
F;3826-3858/Domain: EGF homology <EG15>
F;3866-3909/Domain: LDL receptor YWTD-containing repeat homology <YW40>
F;3910-3968/Domain: LDL receptor YWTD-containing repeat homology <YW41>
F;3943-4543/Domain: alpha-2-macroglobulin receptor 85K chain #status predicted <85K>
F;3943-4420/Domain: 85K chain extracellular #status predicted <EXT>
F;3969-4011/Domain: LDL receptor YWTD-containing repeat homology <YW42>
F;4012-4055/Domain: LDL receptor YWTD-containing repeat homology <YW43>
F;4056-4098/Domain: LDL receptor YWTD-containing repeat homology <YW44>
F;4099-4141/Domain: LDL receptor YWTD-containing repeat homology <YW45>
F;4150-4181/Domain: EGF homology <EG16>
F;4199-4230/Domain: EGF homology <EG17>
F;4235-4266/Domain: EGF homology <EG18>
F;4271-4302/Domain: EGF homology <EG19>
F;4307-4338/Domain: EGF homology <EG20>
F;4343-4373/Domain: EGF homology <EG21>
F;4376-4408/Domain: EGF homology <EG22>
F;4421-4443/Domain: transmembrane #status predicted <TMM>
F;4444-4543/Domain: intracellular #status predicted <INT>
F;116,138,187,276,359,448,731,926,1048,1152,1153,1193,1216,1305,1509,1556,1573,1614,1
3485,3659,3786,3837,3952,4074,4124,4178,4278/Binding site: carbohydrate (Asn) (covale
F;168,2995/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F;2955/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted

```
Alignment Scores:
Pred. No.:              0.511           Length:        4543
Score:                  139.00          Matches:       139
Percent Similarity:     31.52%          Conservative:  69
Best Local Similarity:  21.06%          Mismatches:    196
Query Match:            3.64%           Indels:        256
DB:                     1               Gaps:          37

US-10-008-355-1 (1-2139) x A53102 (1-4543)

Qy    268 TGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGT 327
           |||||| |||:::  :::           ::: :::|||          |||
Db   3765 CysGlyAspGlySerAspGluLysSerCysSerHisGluHisLysSerTyr---AspCys 3783

Qy    328 TTCGTTTCTCGCACGATG---GGTGAGGAGCTTCCG---ATTCCGGGTCTTTCCGTGAAG 381
              :::   ||||||  |||:::|||         |||       |||
Db   3784 MetThrAsnThrThrMetCysGlyAspGluAlaGlnCysIleGlnAlaGlnSerSerThr 3803

Qy    382 TATCTG---------CGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAG 432
              |||         |||       ||||||  ||||||     |||
Db   3804 TyrCysThrCysArgArgGlyPheGlnLysValProAspLysAsnSerCysGln------ 3821

Qy    433 GGTATCACTGACGAGATGGAGCGTCTGCGC------------------------------ 462
                    |||      |||  ||||||
Db   3822 ---------AspValAsnGluCysLeuArgPheGlyThrCysSerGlnLeuCysAsnAsn 3838

Qy    463 ---AAAGCTCAGGAGGTATGC---------CAAGAACTGGCCAAAAAAGAAAAT------ 504
              |||      ||||||       :::         |||   :::|||
Db   3839 ThrLysGlySerHisValCysSerCysAlaLysAsnPheMetLysThrAspAsnMetCys 3858

Qy    505 ------------------------------------GCAGACGAGAACCAACTCTGCATC 528
                                              ||||||:::|||::::::
Db   3859 LysAlaGluGlySerGluHisGlnIleLeuTyrIleAlaAspAspAsnLysIleArgSer 3878

Qy    529 GTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGAC 588
              :::   ||||||      |||:::   |||
Db   3879 MetTyrProPheAsnProAsnSerAlaTyr------------------------------ 3888

Qy    589 GTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGAC--- 645
                                |||:::          |||   ||||||    :::
Db   3889 -----------------GluProAla------------PheGlnGlyAspGluAsnVal 3898

Qy    646 ---------------------------------------------------AACTGGATG 654
                                                             ||||||
Db   3899 ArgIleAspAlaMetAspIleTyrValLysGlyAsnLysIleTyrTrpThrAsnTrp--- 3917

Qy    655 TGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGT------------ 702
                   |||||||||      |||          :::   |||
Db   3918 ---------HisThrGlyArgIleSerTyrCysGluLeuProAlaSerSerAlaAlaSer 3934

Qy    703 ---GCCGACAACCGG--------------------------------------------- 714
              |||  ||||||
Db   3935 ThrAlaSerAsnArgAsnArgArgGlnIleAspGlyGlyValThrHisLeuAsnIleSer 3954

Qy    714 ------------------------------------------------------------ 714

Db   3955 GlyLeuLysMetProArgGlyIleAlaValAspTrpValAlaGlyAsnIleTyrTrpThr 3974
```

```
Qy    715 -----------------------CCGGCCGAATACAGCAAGGACAATAAACCCTATAAG 750
                                    |||:::         :::|||:::
Db   3975 AspSerGlyArgAspValIleGluValAlaGlnMetLysGlyGluAsnArg--------- 3991

Qy    751 CCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACC 810
                             :::   |||          :::   :::|||:::
Db   3992 --------------LysThrLeuIleSerGlyMetIleAspGluProHisAlaIleVal 4006

Qy    811 ATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATC 870
           :::      |||                |||   :::   ||||||
Db   4007 ValAsp---ProLeuArgGlyThrMetTyrTrpSerAspTrpGly-------------- 4020

Qy    871 GAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAA 930
                   |||:::|||:::||||||          :::      |||     :::|||
Db   4021 --------AsnHisProLysIleGluThrAlaAlaMet---AspGlyThrLeuArgGlu 4036

Qy    931 GCCATGAGCGCAGATCAGGCTACCCGTATCAAATAT-----------GCCAGCAAGTAT 978
             :::     |||               |||:::::              |||     |||
Db   4037 ThrLeuValGlnAsp---------AsnIleGlnTrpProThrGlyLeuAlaValAspTyr 4053

Qy    979 GCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTT 1038
                      ||||||   :::                         |||:::|||
Db   4054 HisAsnGluArgLeuTyrTrpAlaAsp-----------------------AlaLysLeu 4065

Qy   1039 GACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAAC 1098
             |||||||||                                   ||||||   |||
Db   4066 SerValIleGlySer-------------------------------IleArgLeuAsn 4074

Qy   1099 GGC--------------------------------------------------------- 1101
          |||
Db   4075 GlyThrAspProValValAlaIleAspAsnLysLysGlyLeuSerHisProPheSerIle 4094

Qy   1102 ---------AAGAGTGCTGTCTATGGC------GATGTATTGTCTTCTCTCGAAAAGGCT 1146
                   :::      :::||||||          :::    :::    :::    |||
Db   4095 AspIlePheGluAspTyrIleTyrGlyValThrTyrIleAsnAsnArgIlePheLysIle 4114

Qy   1147 TATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGT 1206
          :::|||   |||   |||:::          :::|||   |||:::          |||
Db   4115 HisLysPheGlyHisLysSer---------ValThrAsnLeuThrSerGlyLeuAsnHis 4131

Qy   1207 GGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCT 1266
             |||:::||||||   :::   |||:::                 |||||||||
Db   4132 AlaThrAspValValLeuTyrHisGlnTyrLysGlnProGluValThrAsnPro------ 4149

Qy   1267 CATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTC------------ 1314
                                      |||   |||        ::::::|||
Db   4150 --------------------CysAspArgLysLysCysGluTrpLeuCysLeuLeuSer 4162

Qy   1315 CCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGC-------- 1366
          ||||||      |||       ||||||              |||      |||
Db   4163 ProSer-GlyProValCysThrCysProAsnGlyLysArgLeuAspAsnGlyThrCysVa 4182

Qy   1367 ----GTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTC 1422
              :::      |||||||:::|||   ::: :::        :::       ::::::
Db   4182 lLeuIleProSerProThrAlaSerAlaVal-ValProThrThrAspThrCysAspLeuV 4202

Qy   1423 AAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGC 1482
                :::         |||   |||      |||                         :
Db   4202 alCysLeuAsnGlyGlySerCysPheLeuAsnAlaArg--------------------L 4215

Qy   1483 GACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAA---------- 1532
           :::|||      |||      |||            |||:::   ||||||
Db   4215 ysGlnAlaLysCysArgCysGlnProArgTyrAsnGlyGluArgCysGlnIleAsnGlnC 4235

Qy   1533 -----GGCTATCGAGAAAGATCCGGC-----------AGTAGAGCTTTCCAAGAGCGTA 1575
                 |||   :::      |||            |||   :::
Db   4235 ysSerAspTyrCysGlnAsnGlyGlyLeuCysThrAlaSerProSerGlyMetProThrC 4255

Qy   1576 ATAGCTGCTGCTCGCGCTATTCAGGC---CGATGCGATGGCCAA---TGC------CTAT 1623
             |||    :::    ::::::|||   |||||||||   |||   |||          |
Db   4255 ysArgCysProThrGlyPheThrGlySerArgCysAspGlnGlnValCysThrAsnTyrC 4275

Qy   1624 GCCATTGAGAAGGGCAAGCGTCTTTT-----------------CTTTGCCGGTTTGCGT 1665
             |||||      |||          :::                        ||||||
Db   4275 ysHisAsnAsnGlySerCysThrValAsnGlnGlyAsnGlnProAsnCysArg------- 4292

Qy   1666 GAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTAC 1725
                  |||  . |||              :::|||||||||          :::
Db   4293 -----CysProProThrPheIleGlyAspArgCysGlnTyrGlnGlnCysPheAsnTyrC 4311

Qy   1726 GGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGAC 1781
             :::   |||:::                          |||:::|||   |||
Db   4311 ysGluAsnAsnGlyVal--------------------CysGlnMetSerArgAsp 4322
```

RESULT 10
S53362
mucin 5AC (clone JER47) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 08-Oct-1999
C;Accession: S53362; S71065
R;Guyonnet-Duperat, V.; Audie, J.P.; Debailleul, V.; Laine, A.; Buisine, M.P.; Galieg
Biochem. J. 305, 211-219, 1995
A;Title: Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich d
A;Reference number: S53361; MUID:95126907; PMID:7826332
A;Accession: S53362
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-477 <GUY>
A;Cross-references: EMBL:Z34277
A;Experimental source: clone JER47
R;Porchet, N.
submitted to the EMBL Data Library, June 1994
A;Reference number: S71065
A;Accession: S71065
A;Molecule type: mRNA
A;Residues: 1-211,'S',213-224,'AR',227-259,'S',261-477 <POR>
A;Cross-references: EMBL:Z34277; NID:g563374; PIDN:CAA84031.1; PID:g563375
A;Experimental source: clone JER47
C;Genetics:
A;Gene: GDB:MUC5AC
A;Cross-references: GDB:454136; OMIM:158373
A;Map position: 11p15.5-11p15.5
C;Keywords: glycoprotein; tandem repeat

Alignment Scores:

Pred. No.:	0.226	Length:	477
Score:	138.00	Matches:	74
Percent Similarity:	41.18%	Conservative:	31
Best Local Similarity:	29.02%	Mismatches:	94
Query Match:	3.61%	Indels:	56
DB:	2	Gaps:	11

US-10-008-355-1 (1-2139) x S53362 (1-477)

```
Qy   1413 CAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGT 1472
          ::::::|||   ||||||   ||||||:::   ||||||:::   |||   |||
Db     57 GluLysIleCysArgArgProGluGluIleThrArgLeuGlnCysArgAlaGlu------ 74

Qy   1473 TCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAA------GGAAAAGTT 1526
                              |||   :::|||      :::         |||:::|||
Db     75 --------------------SerHisProGluValSerIleGluHisLeuGlyGlnVal 87

Qy   1527 TGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGC 1586
             |||                                                   |||
Db     88 ValGln-------------------------------------------------Cys 90

Qy   1587 TCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCT 1646
          ||||||   |||               :::|||       :::    |||
Db     91 SerArgGluGluGly--------------LeuValCysArgAsnGlnAspGlnGlnGly 105

Qy   1647 TTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCA---- 1702
                   :::|||:::              |||||||||         |||
Db    106 --ProPheLysMetCysLeuAsnTyrGluValArgValLeuCysCysGluThrProLysG 125

Qy   1703 ----------ACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCA 1751
                    ||||||      |||       |||||||||||           |||
```

```
Db    125 lyCysProValThrSerThrProVal------ThrAlaProSerThrProSerGlyArgA 143

Qy   1752 GGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGG---------AGAA 1802
             |||   |||                  |||:::   :::   |||         |||:
Db    143 laThrSerPro---------------ThrGlnSerThrSerSerTrpGlnLysSerArgT 158

Qy   1803 GCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCAC 1862
             ::       |||         |||                 :::|||||||:::|||||||
Db    158 hrThrThrLeuValThrThrSerThrThrSerThrProGlnThrSerThrThrSerAlaP 178

Qy   1863 CAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAA 1922
             ||  |||                    |||  |||:::   |||           :::|
Db    178 roThrThrSerThrThrSerAlaProThrThrSerThrThrSerAlaProThrThrSerT 198

Qy   1923 CGACATCACGGGCGGTAACTCCGGTA------GCCCCG---TATTCGATAAGAACGGCCG 1973
             |||||||       |||.  :::        ||||||      |||        :::|||
Db    198 hrThrSerThrProGlnThrSerIleSerSerAlaProThrThrSerThrThrSerAlaP 218

Qy   1974 TCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTA---------TGAGTGGTGACAT 2024
                |||      :::|||  ||||||        :::                  ::::
Db    218 roThrSerSerThrIleSerArgAlaThrThrSerIleIleSerAlaProThrThrSerT 238

Qy   2025 CGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCAT 2084
             :::::|||   |||       |||     ||||||   |||:::::::|||   |||:::
Db    238 hrThrSerPheProThrThrSerThrThrSerAlaThrThrThrSerThrThrSerAlaP 258

Qy   2085 GATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
                |||:::       |||   |||   :::||||||:::
Db    258 roThrThrSerThrThrSerThrProGlnThrSerLysThr 271
```

RESULT 11
T45025
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T45025
R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A;Reference number: Z22899; MUID:97166151; PMID:9013550
A;Accession: T45025
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3570 <DES>
A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A;Experimental source: placenta
C;Genetics:
A;Gene: MUC5B

```
Alignment Scores:
Pred. No.:              0.631       Length:         3570
Score:                  137.00      Matches:        154
Percent Similarity:     33.29%      Conservative:   94
Best Local Similarity:  20.67%      Mismatches:     293
Query Match:            3.59%       Indels:         204
DB:                     2           Gaps:           33

US-10-008-355-1 (1-2139) x T45025 (1-3570)

Qy     65 CCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAA 124
             |||      |||   :::       :::|||::::::|||   |||
Db   2906 ProThrAlaThrSerSerLysAlaThrSerSerSerSerProArg---------ThrAla 2922

Qy    125 TGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCA 184
                               |||   |||         :::||||||   ::::::        |||
Db   2923 ThrThrLeuProValLeuThrSerThrAlaThrLysSerThrAlaThrSerValThrPro 2942

Qy    185 TTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCG--------------------- 223
             :::|||         |||   :::         :::|||
Db   2943 IleProSerSerThrLeuGlyThrThrGlyThrLeuProGluGlnThrThrThrProVal 2962

Qy    224 GTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCACT---GCGGATACGGTG 280
               :::         |||        |||   |||   ||||||         ||||||
Db   2963 AlaThrMetSerThrIleHisProSerSerThrProGluThrThrHisThrSerThrVal 2982

Qy    281 CTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTC--- 337
          |||:::                         |||:::      :::      |||
Db   2983 LeuThrThrLysAlaThrThrArgAlaThrSerSerThrSerThrProSerSerThrPro 3002

Qy    338 ---GCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGA 394
                   |||:::
Db   3003 GlyThrThrTrpIle--------------------------------------------- 3007

Qy    395 TCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGC 454
                                                            ||||||    :::
Db   3008 ---------------------------------------------LeuThrGluLeuThr 3012

Qy    455 GTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGA 514
                |||                |||       |||                  :::
Db   3013 ThrAlaAlaThrThrThr------AlaGlyThrGlyProThrAlaThrProSerSerThr 3030

Qy    515 ACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACG 574
                :::                 :::    ||||||    |||:::::::::||||||
Db   3031 ProGlyThrThrTrpIleLeuThrGluLeuThrThrThrAlaThrThrThrAlaSerThr 3050

Qy    575 ATGTATTCAAGGACGTTC------------------GTATGGTATTTGCTCCTCCCAGCT 616
                :::   |||                          |||   |||      |||
Db   3051 GlySerThrAlaThrLeuSerSerThrProGlyThrThrTrpIleLeuThrGluPro--- 3069

Qy    617 CTGTAGGTAAGTTCGGAGGCGATACGGACA---------------ACTGGATGTGGCCGC 661
                   |||:::    |||      |||                      |||
Db   3070 ---------SerThrThrAlaThrValThrAlaProProGlySerThrAlaThrAlaSer 3086

Qy    662 GTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCG 721
             |||::::||||||:::                            ||||||
Db   3087 SerThrGlnAlaThrAlaGlyThrProHisValSerThrThrAlaThrThrProThrVal 3106

Qy    722 AATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAG 781
                :::::::::               ||||||   |||:::             |||
Db   3107 ThrSerSerLysAlaThrProSerSerSerProGlyThrAlaThrAlaLeuProAlaLeu 3126

Qy    782 GCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACC 841
                :::      ||||||          |||:::    :::             |||   :::
Db   3127 ArgSerThrAlaThrThr---------ProThrAlaThrSerPheThrAlaIleProSer 3143

Qy    842 TCACTTCTTGGG---GTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAG 898
          |||   ||||||         |||          :::|||:::|||   |||          :::
Db   3144 SerSerLeuGlyThrThrTrpThrArgLeuSerGlnThrThrThrProThrAlaThrMet 3163

Qy    899 TTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTA 958
                      :::      ::::::                            |||
Db   3164 SerThrAlaThrProSerSerThr-----------------------------ProGlu 3173

Qy    959 TCAAATATGCCAGCAAGTATGCTC------AGAGTGCTAACTATTGGAAGAATTCGATCG 1012
          :::         :::::::::|||         |||   |||         |||
Db   3174 ThrValHisThrSerThrValLeuThrThrThrAlaThrThrThrGlyAlaThrGlySer 3193

Qy   1013 GTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCG--------- 1063
          |||   |||      |||         |||               ||||||
Db   3194 ValAlaThrProSerSerThrProGlyThrAlaHisThrThrLysValProThrThrThr 3213

Qy   1064 ---------------AGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTG 1108
                                        |||   |||:::      |||      |||
Db   3214 ThrThrGlyPheThrAlaThrProSerSerSerProGlyThrAlaLeuThrProProVal 3233

Qy   1109 CTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCA 1168
                                 :::|||                            |||
Db   3234 ---------------------TrpIleSerThrThrThrThrProThrThrThrThrPro 3246

Qy   1169 ACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTG 1228
          |||                                 |||:::|||   |||
Db   3247 Thr------------------ThrSerGlySerThrValThrPro--------------- 3255

Qy   1229 CACAGTTTGCCA------ACGCATTGGCTACAAATCCTG--------------------- 1261
             |||:::|||      ||||||      :::::::|||
Db   3256 SerSerIleProGlyThrThrHisThrAlaArgValLeuThrThrThrThrThrThrVal 3275
```

```
Qy    1262 ------------ATGCTCATGCCGGTATCCTCAAATCGC--------------------- 1288
                       |||       |||  ||||||   :::
Db    3276 AlaThrGlySerMetAlaThrProSerSerSerThrGlnThrSerGlyThrProProSer 3295

Qy    1289 ---------------------TTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACC 1327
                                :::|||   :::      ||||||:::|||   ||||||
Db    3296 LeuThrThrThrAlaThrThrIleThrAlaThrGlySerThrThrAsnProSerSerThr 3315

Qy    1328 GTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCG------ACA 1381
                       |||         :::|||  :::       :::   |||      |||
Db    3316 ProGlyThrThrProIleProProValLeuThrSerMetAlaThrThrProAlaAlaThr 3335

Qy    1382 AGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGT 1441
           ||||||         |||         |||:::            ||||||
Db    3336 SerSerLysAlaThrSer---------SerSerSerProArgThrAlaThrThrLeuPro 3352

Qy    1442 ATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGT------TCCATG 1495
           :::   ||||||   :::|||:::               |||||||||         :::
Db    3353 ValLeuThrSerThrAlaThrLys---------SerThrAlaThrSerPheThrProIle 3369

Qy    1496 CCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAG 1555
           |||   |||:::   ||||||
Db    3370 ProSerSerThrLeuTrpThr--------------------------------------- 3376

Qy    1556 TAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCA 1615
                                                                  |||
Db    3377 -------------------------------------------------------ThrTrpThr 3379

Qy    1616 ATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACC 1675
           :::|||                                    :::              |||
Db    3380 ValProAlaGlnThrThrThrProMetSerThrMetSerThrIleHisThrSerSerThr 3399

Qy    1676 CCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCA--- 1732
           |||:::                     ||||||     :::       ||||||
Db    3400 ProGlu-----------ThrThrHisThrSerThrValLeuThrThrThrAlaThrMet 3415

Qy    1733 TCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCG 1792
           :::|||      |||            |||   :::|||:::
Db    3416 ThrArgAlaThrAsnSerThrAlaThrProSerSerThrLeu------------------ 3429

Qy    1793 TATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACC 1852
                    :::|||||||||            ::::::                   :::::::|||
Db    3430 ---GlyThrThrArgIleLeuThrGluLeuThrThrThrAlaThrThrThrAlaAlaThr 3448

Qy    1853 ------------------TCTTCCGCACCAAAAACT------ATGGTCGCTATGCCGAGA 1888
                             ||||||   |||   |||      ::::::       |||
Db    3449 GlySerThrAlaThrLeuSerSerThrProGlyThrThrTrpIleLeuThrGluProSer 3468

Qy    1889 ACGGTCAGCTCCATATCGCTT---------TCCTATCGAACAACGACATCACGGGCGGTA 1939
           |||::::::::::::   :::         |||      ||||||:::|||
Db    3469 ThrIleAlaThrValMetValProThrGlySerThrAlaThrThrSerSerThrLeuGly 3488

Qy    1940 ACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCA 1999
           |||         |||                              |||:::      :::||||||
Db    3489 ThrAlaHisThrPro------------------------LysValValThrAlaMetAla 3500

Qy    2000 ACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCG 2059
           |||            :::   |||::::::::::::   :::   |||:::   :::
Db    3501 Thr-----------MetProThrAlaThrAlaSerThrValProSerSerSerThrVal 3516

Qy    2060 TGGACATCCGCTACG 2074
              |||:::   |||
Db    3517 GlyThrThrArgThr 3521
```

RESULT 12
A60999
alpha-amylase (EC 3.2.1.1) precursor - Micrococcus sp. (strain 207)
C;Species: Micrococcus sp.
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999
C;Accession: A60999
R;Kimura, T.; Horikoshi, K.
FEMS Microbiol. Lett. 71, 35-42, 1990
A;Title: The nucleotide sequence of an alpha-amylase gene from an alkalopsychrotrophi
A;Reference number: A60999
A;Accession: A60999
A;Molecule type: DNA
A;Residues: 1-1104 <KIM>
A;Cross-references: GB:X55799; NID:g296762; PIDN:CAA39321.1; PID:g296763
C;Function:
A;Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A;Pathway: glycogen/starch degradation
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-1104/Product: alpha-amylase #status predicted <MAT>

```
Alignment Scores:
Pred. No.:               0.659        Length:        1104
Score:                   133.50       Matches:       138
Percent Similarity:      34.94%       Conservative:  94
Best Local Similarity:   20.78%       Mismatches:    259
Query Match:             3.49%        Indels:        173
DB:                      2            Gaps:          35
```

US-10-008-355-1 (1-2139) x A60999 (1-1104)

```
Qy     421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480
           |||      |||||||||||||||:::::::::            |||
Db     384 GlyAspPheLysGlyIleThrAspLysLeuAspTyrLeu--------------------- 396

Qy     481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540
                             :::              |||   :::   |||      |||
Db     397 ------------------AspGluLeuGlyValAsnThrIleTrpIle---SerProVal 409

Qy     541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600
               |||     :::|||                              ||||||||
Db     410 ValGluAsnIleLysTyr---------------------------AspValArgTyrTyr 420

Qy     601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGG------ATG 654
                                |||      |||         :::|||:::      :::
Db     421 GluThrSerGluProTyrTyrGlyTyrHisGlyTyrTrpAlaAsnAsnPheGlyGluLeu 440

Qy     655 TGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGC---GTGTATGCCGGTGCCGACAAC 711
            |||   |||   |||                |||      :::      ||||||      :::
Db     441 AsnPro---HisPheGlyThrMetGluGluPheHisAspLeuIleAspGlyAlaHisAsp 459

Qy     712 CGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTA 771
           |||      :::         |||                                    |||
Db     460 ArgAsnMetLysIleMetValAsp---------------------------ValValVal 470

Qy     772 TCCATGCAAGGCTAC------AAGGCTGACGACTATGCCATGACC---------ATCGGT 816
           :::      ||||||      |||      |||   :::::::|||         |||
Db     471 AsnHisThrGlyTyrGlyLeuLysGluIleAspGlySerValThrAsnProProAlaGly 490

Qy     817 TTCCCGGGCAGTACGGATCGCTACCTCACTTCT--------------------------- 849
           :::|||         ||||||         |||
Db     491 TyrProSerAspAlaAspArgAlaArgPheSerAspLeuLeuArgGlnGlyAlaAspVal 510

Qy     850 ---------------------------TGGGGTGTGGAAGATCGTATCGAAAACGAGAAC 882
                                         |||:::   |||   |||      :::
Db     511 GlyThrAspGluValValGlyGluLeuAlaGlyLeuProAspPheIleThrGluAspPro 530

Qy     883 AATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCA 942
           |||   |||   ::::::                  |||      |||   |||      :::
Db     531 AsnValArgLysGlnIleIleAsp---TrpGlnThrAspTrpIleGluLysAlaThrThr 549

Qy     943 GATCAGGCTACCCGTATCAAATAT------------GCCAGCAAGTATGCTCAGAGTGCT 990
           :::            |||   |||                     |||:::   :::   |||
Db     550 GluAsnGlyAsnThrIleAspTyrPheArgValAspThrValLysHisValGluAspAla 569

Qy     991 AACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGT 1050
              :::   :::            :::         :::         :::||||||
Db     570 ThrTrpMetGlnPheLysAsnAlaLeuThrGluLysMetProGluPheLysMetIleGly 589

Qy    1051 CGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCT 1110
                                            |||      |||         :::
```

```
Db     590 GluAla--------------------------TrpGlyAlaLysValAspAsnThrLeu 600

Qy    1111 GTCTAT---------GGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCC 1161
              |||          |||   :::   |||  |||:::       :::||||||   |||
Db     601 GlyTyrLeuGluThrGlyThrMetAspSerLeuLeuAspPheGlyPheLysGluThrAla 620

Qy    1162 AAG--------------------------GCCAACCGTGAGATGACTTATTTGAGCGAG 1194
           :::                           ||||||       :::|||     :::
Db     621 ArgSerPheValAsnGlySerLeuGluAlaAlaAsnAlaSerLeuThrAlaArgAsnAla 640

Qy    1195 ACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTT------------------GCACAGTTT 1236
              |||             :::   :::|||                        |||
Db     641 LysLeuAspAsnThrAlaThrLeuGlyGlnPheLeuGlySerHisAspGluGluGlyPhe 660

Qy    1237 GCCAACGCATTGGCTACAAAT--------------------------------------- 1257
              :::::||||||   :::
Db     661 LeuHisSerLeuAlaGlyAspLysGlyLysLeuGlnValAlaAlaThrLeuGlnAlaThr 680

Qy    1258 ------------CCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAA 1305
                       |||      :::   |||       :::   |||
Db     681 AlaLysGlyGlnProValIleTyrTyrGly------GluGluLeuGlyGlnThrGlyAla 698

Qy    1306 GACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGG 1365
           :::         |||      |||   :::                          |||
Db     699 AsnAsnTyrProGlnTyrAspAsnArg---------------TyrAspPheAlaTrpAsp 713

Qy    1366 CGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAA 1425
           ::::::      :::::::::          :::      :::::::::   :::   :::
Db     714 GlnValGluGlyAsnGluIleLeuAlaHisTyrThrLysIleLeuAsnPheArg---Glu 732

Qy    1426 GGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGAC 1485
           |||   :::|||   :::|||             :::::::::::::|||     ||||||
Db     733 GlyTyrSerLysValPheAlaLys------GlyGluArgThrLeuValGlyGlySerAsp 750

Qy    1486 AAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAA 1545
           |||         :::|||   |||   |||   :::                :::
Db     751 LysAspGlnPheLeuLeuPheSerArgAspTyrGlnAspGlnLysValTyrValGlyLeu 770

Qy    1546 GATCCGGCAGTAGAGCTTTCCAAGAGCGTA---------ATAGCTGCTGCTCGCGCTATT 1596
           :::   |||   |||   ||||||:::|||         :::   :::|||   |||:::
Db     771 AsnValAlaGluGlu---SerLysAlaValThrLeuThrValAspSerAlaAspAlaVal 789

Qy    1597 CAGGCCGATGCGATGGCCAATGCC---TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTT 1653
              ||||||   :::            |||         ||||||
Db     790 ValThrAspAlaTyrSerGlyThrGluTyrThrAlaThrAlaGlyLys------------ 805

Qy    1654 GCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATG 1713
                 :::         ||||||:::|||   |||                 |||
Db     806 ------ValAsnLeuThrLeuProGlyLysAlaAspProGlyThrValLeuLeuThr--- 822

Qy    1714 CGTATGAGCTACGGCTCCATCAAGGGA--------------------------------- 1740
              :::   |||:::|||   |||
Db     823 ---ValGluGlyGlyAsnIleThrGlyValAlaLysAspAsnGlyGluValValValGlu 841

Qy    1741 -----------------------------TATGAACCGCAGGACGGTGCCTGGTACAAC 1770
                                        |||:::   :::|||      :::   |||
Db     842 LeuValProGluAsnAsnIleArgIleHisTyrLysArgGluAspAsnValTyrLysAsn 861

Qy    1771 TATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCC 1830
           |||          |||   |||   :::                  |||    |||
Db     862 Tyr---------GlyAlaTrpLeuTrpAsnAspValAlaSerProSerAlaAsnTrpPro 878

Qy    1831 GTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAG--- 1887
           |||         :::   :::         :::|||   :::||||||   |||   :::
Db     879 ValGlyAlaThrMetPheGlu------LysThrAspSerTyrGlyAlaTyrIleAspVal 896

Qy    1888 ---------AACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGT 1938
                    :::|||      :::|||   ||||||         |||||||||   |||
Db     897 ProLeuAlaAspGlyAlaLysAsnIleGlyPheLeuVal---MetAspIleThrAlaGly 915

Qy    1939 AACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGC 1998
           ::::::|||                  ||||||         :::               :::
Db     916 AspAlaGlyLysAspGlyGlyAspLysGlyPheThrIleSerSerProGlnAlaAsnGlu 935

Qy    1999 AACTGGGAAGCTATGAGTGGTGAC------ATCGAGTTCGAACCC---GATCTGCAGCGC 2049
                 |||               |||            :::||||||   ||||||
Db,    936 IleTrpIleLysGlnGlySerAspLysValTyrThrTyrGluProValAspLeuProAla 955

Qy    2050 ACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGAC------------------ 2091
                 :::|||   |||   |||               :::|||
Db     956 Asn---ThrValArgIleHisTyrThrArgGluAlaValAspTyrAspAspPheGlyIle 974

Qy    2092 ---AAATGGGGT 2100
                 ||||||
Db     975 TrpAsnTrpGly 978
```

RESULT 13
D97933
valine-tRNA ligase (EC 6.1.1.9) [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 03-Jun-2002
C;Accession: D97933
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: D97933
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-883 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAK99296.1; PID:g15458063; GSPDB:GN00174
C;Genetics:
A;Gene: valS
C;Superfamily: valine-tRNA ligase
C;Keywords: ligase

```
Alignment Scores:
Pred. No.:                0.647          Length:         883
Score:                    133.00         Matches:        104
Percent Similarity:       33.07%         Conservative:   65
Best Local Similarity:    20.35%         Mismatches:     166
Query Match:              3.48%          Indels:         176
DB:                       2              Gaps:           24
```

US-10-008-355-1 (1-2139) x D97933 (1-883)

```
Qy     574 GATGTATTC------------------------AAGGACGTTCGTATGGTATTTGCTCCT 609
           |||||||||                        |||         :::|||      |||
Db      26 AspValPheLysProSerGlyAspGlnLysAlaLysProTyrSerIleValIleProPro 45

Qy     610 CCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACG 669
           |||:::      ||||||                                       |||
Db      46 ProAsnValThrGlyLysLeu---------------------------------HisLeu 54

Qy     670 GGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGC 729
           |||                                                   ::::::
Db      55 Gly---------------------------------------------------HisAla 57

Qy     730 AAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAG 789
           |||         :::    :::                  |||||||||:::
Db      58 TrpAspThr---ThrLeuGlnAspIleIleIleArgGlnLysArgMetGlnGlyPheAsp 76

Qy     790 GCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTC---ACT 846
                             |||:::      ||||||               :::   |||
Db      77 -----------------ThrLeuTrpLeuProGlyMetAspHisAlaGlyIleAlaThr 90

Qy     847 TCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGT 906
                    ||||||:::|||:::         |||         |||   ::::::
Db      91 GlnAlaLysValGluGluArgLeuArgGlyGluGlyIleSerArgTyrAspLeuGlyArg 110

Qy     907 ------------ATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACC 954
                       |||         ||||||:::                :::   ||||||
Db     111 GluSerPheLeuThrLysValTrpGluTrpLysAsp------------GluTyrAlaThr 126
```

```
Qy    955 CGTATCAAATATGCC---AGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATC 1011
              ||||||             |||         |||  :::|||   :::
Db    127 ThrIleLysGluGlnTrpGlyLysMetGlyLeuSerValAspTyrSerArgGluArgPhe 146

Qy   1012 GGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGA 1071
             ::::::  ||||||:::                           :::|||       :::
Db    147 ThrLeuAspGluGlyLeuSer-----------------------LysAlaValArgLys 158

Qy   1072 GCATTCGCAGACTGGATCCGTAAG------------AACGGCAAGAGTGCTGTCTATGGC 1119
             |||   |||       :::|||              |||:::        :::
Db    159 ValPheValAspLeuTyrLysLysGlyTrpIleTyrArgGlyGluPheIleIleAsnTrp 178

Qy   1120 GAT---------------GTATTGTCTTCTCTCGAAAAGGCTTATAAG------------ 1152
          |||                |||||    :::|||      :::|||
Db    179 AspProAlaAlaArgThrAlaLeuSerAspIleGluValIleHisLysAspValGluGly 198

Qy   1153 -----------------------------GAAGGAGCCAAGGCCAACCGTGAGATGACT 1182
                                       :::|||:::::::|||             |||
Db    199 AlaPheTyrHisMetAsnTyrMetLeuGluAspGlySerArgAlaLeuGluValAlaThr 218

Qy   1183 TATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAAC 1242
                   ||||||:::||||||
Db    219 ThrArgProGluThrMetPheGlyAspVal------------------------------ 228

Qy   1243 GCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTAC 1302
          |||:::|||   ||||||:::                           |||   :::|||
Db    229 AlaIleAlaValAsnProGlu---------------------------AspProArgTyr 239

Qy   1303 AAAGACTACCTCCCCTCG---------------CTCGACCGTAAGGTGCTGCCCGCCATG 1347
          ||||||   :::                    :::   |||::::::|||    :::
Db    240 LysAspLeuIleGlyLysAsnValIleLeuProIleAlaAsnLysLeuIleProIleVal 259

Qy   1348 CTCGAT---------------------------------------ATTGTACGCCGGCGT 1368
             |||                                           |||
Db    260 GlyAspGluHisAlaAspProGluPheGlyThrGlyValValLysIleThrProAlaHis 279

Qy   1369 ATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC 1428
             |||   |||   |||                |||:::
Db    280 AspProAsnAspPheLeuValGlyGlnArgHisAsnLeuProGlnValAsnIleMetAsn 299

Qy   1429 GACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAG 1488
          |||                :::   ||||||:::   |||   :::        |||:::
Db    300 AspAspGlyThrMetAsnGluLeuValPheGluPheSerGlyMet---------AspArg 316

Qy   1489 TTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGAT 1548
          |||   |||                     |||       ||||||      |||
Db    317 PheGluAlaArg---------------LysAlaValValAlaLysLeuGluGluIleGly 331

Qy   1549 CCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCT--------------- 1593
                |||::::::   |||   |||     :::          :::
Db    332 AlaLeuValLysIleGluLysArgValHisSerValGlyHisSerGluArgThrGlyVal 351

Qy   1594 ------------------------------------------------ATTCAGGCCGAT 1605
                                                          :::      :::
Db    352 ValValGluProArgLeuSerThrGlnTrpPheValLysMetAspGlnLeuAlaLysAsn 371

Qy   1606 GCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGT 1665
          |||:::||||||               :::|||
Db    372 AlaIleAlaAsnGlnAspThrGluAspLys---------------------------Val 382

Qy   1666 GAGATGTACCCCGGACGTGCTCTGCCGAGC---------------------------GAT 1698
          |||   ||||||   |||            :::                        |||
Db    383 GluPheTyrProProArgPheAsnAspThrPheLeuGlnTrpMetGluAsnValHisAsp 402

Qy   1699 GCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGT 1758
             :::   ::::::   :::|||                        |||
Db    403 TrpValIleSerArgGlnLeuTrpTrpGlyHis------------------GlnIlePro 416

Qy   1759 GCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGC 1818
          ||||||||||||          |||:::      :::   ::::::   |||:::
Db    417 AlaTrpTyrAsn------AlaAspGlyGluMetTyrValGlyGluGluAlaProGluGly 434

Qy   1819 GATGAGTTTGCCGTACAGGAGAATATCCTCGAC 1851
          |||   :::         |||::::::||||||
Db    435 AspGlyTrpThrGlnAspGluAspValLeuAsp 445
```

RESULT 14
C71528
probable pts pep phosphotransferase - Chlamydia trachomatis (serotype D, strain UW3/C
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Jan-2000
C;Accession: C71528
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A;Reference number: A71570; MUID:99000809; PMID:9784136
A;Accession: C71528
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-571 <ARN>
A;Cross-references: GB:AE001306; GB:AE001273; NID:g3328748; PIDN:AAC67931.1; PID:g332
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: ptsI
C;Superfamily: phosphotransferase system enzyme I; phosphotransferase system enzyme I
F;17-565/Domain: phosphotransferase system enzyme I homology <PTl>

```
Alignment Scores:
Pred. No.:              0.628        Length:        571
Score:                  132.00       Matches:       115
Percent Similarity:     35.37%       Conservative:  88
Best Local Similarity:  20.03%       Mismatches:    186
Query Match:            3.33%        Indels:        185
DB:                     2            Gaps:          24

US-10-008-355-1 (1-2139) x C71528 (1-571)

Qy   1635 CTTCTCAATGCATAGGCATTGGCCATCGCATCGGCCTGAATAGCGCGAGCAGCAGCTAT 1576
          ::::::   |||||||||:::|||       :::
Db     22 IleValProGlyIleGlyLeuGlyLysAlaLeu-------------------------- 32

Qy   1575 TACGCTCTTGGAAAGCTCTACTGCCGGATCTTTCTCGATAGCCTTGGCAAACTTTTCCTT 1516
             |||||||||       ||||||       :::|||   :::      :::
Db     33 ---LeuLeuGlyLysSerSerLeuArgIleArgGluLeuThrLeuProGlnGluGluVal 51

Qy   1515 GTCCATGGACTTGAGCAT-----------------------GGCATGGAACTTGTCGCT 1480
             |||   :::                                :::|||   |||
Db     52 GluHisGluIleSerArgTyrTyrLysAlaLeuLysArgSerArgSerAspLeuAlaAla 71

Qy   1479 ATAAGGAACCACACTCTTGTCGAA---------TACGAAGTCTGCATACTTCTTCGTGTC 1429
          :::                  :::         |||:::      :::|||:::
Db     72 LeuGluLysGluAlaLysGlyLysGlnGlyTyrGlnGluIleAlaSerIleLeuGlnAla 91

Qy   1428 GCCTTTGAATTTCTTGTCGATTACATTCTT---GAATATATCGGGGAGCTTGTCGGCAGG 1372
             |||   :::   |||   :::|||   |||   :::   :::   :::   :::
Db     92 HisLeuGluIleIleLysAspProLeuLeuThrGluGluValValLysThrIleArgLys 111

Qy   1371 GATACGCCGGCGTACAATATCGAGCATGGCGGGCAGCACCTTACGGTCGAGCGAGGGGAG 1312
          |||                                          :::   |||
Db    112 AspArg------------------------------------------LysAsnAlaGlu 117

Qy   1311 GTAGTCTTTGTACTTGTCGTCAAGCGATTTGAGGATACCGGCATGAGCATCAGG------ 1258
             ||||||   |||:::         |||   :::      ::::::::
Db    118 PheValPheSerSerValMetGlyGluIleGluLysSerLeuCysAlaValGlnLysThr 137

Qy   1257 ---------------------------------------ATTTGTAGCCAATGCGTTGGC 1237
                                                 |||   ::::::   :::|||
Db    138 ThrAlaThrArgValAspArgValGlnAspIleHisAspIleSerAsnArgValIleGly 157

Qy   1236 AAACTGTGC------AAAACGAACCACCTCGGTACCACCGAAGAGCGTCTC--------- 1192
             ||||||         :::   ||||||      ::::::   |||
Db    158 HisLeuCysCysGlnHisLysSerSerLeuGlyGluPheAspGlnAsnLeuIleValPhe 177

Qy   1192 ------------------------------------------------------------ 1192
```

```
Db    178 SerGluGluLeuThrProSerGluAlaAlaAsnAlaAsnProGluTyrIleArgGlyPhe 197

Qy   1191 --------------GCTCAAATAAGTCATCTCACGGTTGGCCTTGGCTCCTTCCTTATA 1147
                        |||:::  ||||||      :::                   :::
Db    198 ValSerLeuGluGlyAlaLysThrSerHisThrAlaIle-----------------Val 211

Qy   1146 AGCCTTTTCGAGAGAAGACAATACATCGCC----------------------------- 1117
          ||||||   :::          |||:::|||
Db    212 SerLeuAlaLysAsnIleProTyrValAlaAsnPheThrThrGluLeuTrpAspThrIle 231

Qy   1116 ------ATAGACAGCACTCTTGCCGTTCTTACGGATCCAGTCTGCGAATGCTCTTTC--- 1066
                        ||||||  ::::::                 |||   :::|||
Db    232 LysGluPheSerGlyThrLeuValLeuIleAsnGlyAspLysGlyGluIleThrPheAsn 251

Qy   1065 ------CTCGGCACGCTTACGACCTATCACGTCAAG-----------------ACGAGC 1030
                |||  |||:::  |||||||:::  |||                 |||
Db    252 ProGlnLeuSerThrIleGlnThrTyrTyrArgLysGlnAlaSerValSerValThrVal 271

Qy   1029 GAGACCGCGGTTCATACCGATCGAATTCTTCCAATAGTTAGCACTCTGAGCATACTTGCT 970
                   |||  |||   :::  |||||||:::::::||| |||   |||
Db    272 ProValGlnValGlnThrGlyLysAsnLeuProLeuIleSer-LeuSerAla-------- 288

Qy    969 GGCATATTTGATACGGGTAGCCTGATCTGCGCTCATGGCTTCCTTCCAGATGCCTTGCTT 910
                                      ::::::::::::  ::::::|||   ::
Db    289 ---------------------------GlnIleValSerThrGluGluLeuProMetIl 299

Qy    909 G---------ATACCGCGAACTTCGATACGAGGATTGTTCTCGTTTTCGATACGATCTTC 859
          :         |||  |||||||:::   ||||||||
Db    299 eGluArgGluSerProGlyThrSerVal---GlyLeuPhe------------------- 311

Qy    858 CACACCCCAAGAAGTGAGGTAGCGATCCGTACTGCCCGGGAAACCGATGGTCATGGCATA 799
                                                  :::         ||||||
Db    312 ---------------------------------------ArgSerGluPheMetAlaPh 318

Qy    798 GTCGTCAGCCTTGTAGCCTTGCATGGATACGGCAGCGAAGTAAACGGGCTTATAGGGTTT 739
           |||              ||||||::::::  |||      :::   |||     |||
Db    318 eSerLeuGlyArgLeuProCysValGlu-GluGlnAlaAspGlnTyrAlaGlnLeuValG 338

Qy    738 AT---------------------------------------------------------T 736
                                                                     |
Db    338 lnPheGlnCysSerAspIleHisValLeuArgLeuPheAspPheGlyGluAspLysGluC 358

Qy    735 GTCCTTGCTGTATTCGGCCGGCCGGTTGTCGGCACCGGCATACACGCGGAATACGCTGAA 676
          ||||||||                  ||||||||    |||   :::    |||
Db    358 ysProCys------------IleSerSerSerHisArgSerValArgTrpLeu-LeuGlu 373

Qy    675 GTCGCCCGTGTGACGCGGCCACAT---CCAGTTGTCCGTATCGCCTCCGAACTTACCTAC 619
                :::|||   :::              :::   |||   |||
Db    374 GlnGluLysValLeuLysGluGlnLeuGlnAlaIleAlaIleValSer----------- 389

Qy    618 AGAGCTGGGAGGAGCAAATACCATACGAACGTCCTTGAATACATCGTAGACGATGAGGAA 559
          |||  ||||||  |||               ::::::     :::::::|||     |||
Db    390 ArgIleGlyArgLeuLysVal---------LeuIleProGlyValIleAspAlaSerGlu 406

Qy    558 GTATTCGTTGTTGGAATAGAAAGGCTCTACGATGCAGAGTTGGTTCTCGTCTGCATTTTC 499
          :::   :::|||      :::||||||:::      |||:::   |||:::
Db    407 IleAlaLeuVal------LysArgLeuPheGlnGluGluIleArgLeuLeuLysGlyIle 424

Qy    498 TTTTTGGCCAGTTCTTGGCATACCTCCTGAGCTTTGCGCAGACGCTCCATCTCGTCAGT 439
                                 ::::::           :::   ::::::
Db    425 SerGluAsnIleLeuTrpGlySerMetIleGluIleProSerAlaValTrpMetIleGlu 444

Qy    438 GATACCCTTGAGCTGTCCTTCTACCTTGTCCGTTACCTTCACGATCTTGCGCAGATACTT 379
          :::   |||:::   |||     ::::::            ::::||||||||||||
Db    445 GluIleLeuGlnGluSerSerPheValAlaLeuGlyThrAsnAspLeuAlaGlnTyrThr 464

Qy    378 CACGGAAAGACCCGGAATCGGAAGCTCCTCACCCATCGTGCGAGAAACGAAACCATCGCG 319
             |||   :::|||   |||   ||||||   |||:::|||              :::
Db    465 LeuGlyThrSerArgGluArgSerLeuLeuGlyGluArgSerArgValProHisProSer 484

Qy    318 CAGATAGTCGTGATCCACCGTGCTTTG------GCTCTGGATAGCACCGTATCCGCAGTG 265
             |||   :::||||||      :::         |||         |||
Db    485 ValIleArgMetIleHisHisValValGluGlnAlaLysGlnLysAsnValProValSer 504

Qy    264 GTGGTTGGTAAA---------GATCAGGCCCTGATC 238
           |||   |||:::          |||   ||||||:::
Db    505 ValCysGlyGluMetAlaGlyAspProAlaLeuLeu 516
```

RESULT 15
PC4395
mucin 3 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 10-Nov-1997 #sequence_revision 10-Nov-1997 #text_change 08-Oct-1999
C;Accession: PC4395
R;Van Klinken, B.J.W.; Van Dijken, T.C.; Oussoren, E.; Bueller, H.A.; Dekker, J.; Ein
Biochem. Biophys. Res. Commun. 238, 143-148, 1997
A;Title: Molecular cloning of human MUC3 cDNA reveals a novel 59 amino acid tandem re
A;Reference number: PC4395; MUID:97445141; PMID:9299468
A;Accession: PC4395
A;Molecule type: mRNA
A;Residues: 1-648 <VAN>
A;Cross-references: DDBJ:AF016692; NID:g2454614; PIDN:AAB71685.1; PID:g2454615
A;Experimental source: intestine
F;1-59,60-118,119-177,178-236,237-295/Region: repeat
F;296-565/Region: semi-unique #status predicted
F;566-582,583-599,600-616,617-633,634-647/Region: repeat

Alignment Scores:

Pred. No.:	0.776	Length:	648
Score:	131.00	Matches:	140
Percent Similarity:	34.14%	Conservative:	71
Best Local Similarity:	22.65%	Mismatches:	173
Query Match:	3.43%	Indels:	234
DB:	2	Gaps:	34

US-10-008-355-1 (1-2139) x PC4395 (1-648)

```
Qy    536 CTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATG---TATTCAAGGACGTTC 592
             :::|||   |||:::      :::||||||   :::||||||      |||   |||:::
Db    104 MetSerThrProSerGluGlySerSerSerLeuThrThrMetLeuLeuSerSerThrTyr 123

Qy    593 GTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATA------------ 640
          |||                            ||||||||||||
Db    124 Val---------------------------ThrSerSerGluAlaSerThrProSerThr 134

Qy    641 -----------CGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCC 688
                        |||:::|||            ||||||   :::|||   :::   |||
Db    135 ProSerValAspArgSerThr---------ProValThrThrSerThrGlnSerAsnSer 151

Qy    689 GCG--------------------TGTATGCCGGTGCCGACAACCGGCCGGCCGAATACA 727
                                 :::|||:::   |||            :::|||
Db    152 ThrProThrProProGluValIleThrLeuProMetSerThrProSerGluValSerThr 171

Qy    728 GCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACA 787
                ||||||   |||:::|||      |||:::                  :::|||
Db    172 ProLeuThrIleMetProValSer------ThrThrSerValThrIleSerGluAlaGly 189

Qy    788 AGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCA--- 844
                :::|||:::|||            |||      :::      :::   ||||||
Db    190 ThrAlaSerThrLeuPro---------ValAspThrSerThrProValIleThrSerThr 206

Qy    845 -----------------------------------CTTCTTGGGGTGTGGAAGATC--- 865
                                              :::   :::|||   :::
Db    207 GlnValSerSerSerProValThrProGluGlyThrThrMetProIleTrpThrLeuSer 226

Qy    866 ----------------------------GTATCGAAAACGAGAACA------------ 883
                                      ||||||   ||||||
Db    227 GluGlySerIleProIleThrIleMetHisValSerThrThrArgValThrSerSerGlu 246

Qy    884 ----------------------------------------ATCCTCGTATCGAAGTTC 901
                                                  :::      |||
Db    247 GlySerThrLeuSerThrProSerValValThrSerThrProValThrThrSerThrGlu 266

Qy    902 GCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCA 961
          |||:::||||||   |||:::                          |||      :::
Db    267 AlaIleSerSerSerAlaThr--------------------------LeuAspSerThr 277
```

```
Qy     962 AATATGCCAGCAAGTATG---CTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGA 1018
              |||      ||||||   :::   :::   |||:::|||
Db     278 ThrMetSerValSerMetProMetGluIleSerThrLeuGly----------------- 291

Qy    1019 ACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCG 1078
                          ||||||  |||      |||         |||
Db     292 ---ThrThrIleLeuValSerThrThrProValThrArgPheProGluSerSerThrPro 310

Qy    1079 CAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCG 1138
                        ||||||  |||               ||||||:::|||
Db     311 SerIleProSerValTyrThr-----------SerMetSerMetThrThr--------- 323

Qy    1139 AAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGC 1198
                                                                |||
Db     324 -------------------------------------------------AlaSerGlu 326

Qy    1199 TCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATC 1258
              |||                          |||  ||||||   |||:::
Db     327 GlySer-------------------------SerSerProThr---ThrLeuGluGly 336

Qy    1259 CTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCT 1318
                        ||||||:::|||                         |||||||||
Db     337 ThrThrThrMetProMetSer-------------------------ThrThrSerGlu 347

Qy    1319 CGCTCGACCGTAAGG---------TGCTGCCCGCCATGCTCGATATTG----------- 1357
           |||||||||:::                    |||   |||::::::
Db     348 ArgSerThrLeuLeuThrThrValLeuIleSerProIleSerValMetSerProSerGlu 367

Qy    1358 TACGCCGGCGTATCCCTGCCG---------ACAAGCTCCCCGATATATTCAAGAATGTAA 1408
              :::   :::|||  |||          ||||||:::|||:::    :::
Db     368 AlaSerThrLeuSerThrProProGlyAspThrSerThrProLeuLeuThr--------- 384

Qy    1409 TCGACAAGAAATTCAAAG----------------------GCGACACGAAGAAGTATG 1444
           ||||||:::                                   |||   |||:::
Db     385 SerThrLysAlaGlySerPheSerIleProAlaGluValThrThrIleArgIleSerIle 404

Qy    1445 CAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCA 1504
              :::      ||||||  :::      |||:::  ::::::::::::|||   |||
Db     405 ThrSerGluArgSerThrProLeuThrThrLeuLeuValSerThrThrLeuProThrSer 424

Qy    1505 AGTCCATGGACAAGG---AAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGC 1561
              |||                  |||   |||
Db     425 PheProGlyAlaSerIleAlaSerThrPro----------------------------- 434

Qy    1562 TTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCT 1621

Db     434 ------------------------------------------------------------ 434

Qy    1622 ATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGAC 1681
              ||||||     ::::::   |||                        ||||||
Db     435 ---ProLeuAspThrSerThrThrPhe-----------------------ThrProSer 445

Qy    1682 GTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGAT 1741
                        :::   ||||||   |||   |||      |||
Db     446 ThrAspThrAlaSerThrProThrIlePro---ValAlaThrThrIleSerValSerVal 464

Qy    1742 ATGAACCGCAGGACGGTGCCTGGTACAACTATC-----------------ATACGACAG 1783
           :::      :::   |||||||||||||||                  :::
Db     465 IleThrGluGlySerThrProGlyThrThrIlePheIleProSerThrProValThrSer 484

Qy    1784 GCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATA 1843
           :::   |||                     |||:::|||   |||          |||
Db     485 SerThrAlaAspValPheProAlaThrThrGlyAlaValSerThrProVal------Ile 502

Qy    1844 TCCTCGACCTCT---TCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCC 1900
           :::||||||      :::   |||   |||                   :::   ||||||
Db     503 ThrSerThrGluLeuAsnThrProSerThr------------------SerSerSerSer 516

Qy    1901 ATATCGCTTTCCTATCGAACAACG---------ACATCACGGGCGGTAACTCCGGTAGCC 1951
              :::   |||:::   ||||||          |||:::   |||:::|||      |||
Db     517 ThrThrThrSerPheSerThrThrLysGluPheThrThrProAlaMetThrThrAlaAla 536

Qy    1952 CCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTA 2011
           |||   :::   |||
Db     537 ProLeuThrTyrValThr------------------------------------------ 542

Qy    2012 TGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCT 2071
              :::   |||:::   |||   |||      ::::::::            |||||||||
Db     543 ---MetSerThrAlaProSerThrProArgThrThrSerArgGlyCysThrThrSerAla 561

Qy    2072 ACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
           :::      |||                     |||   |||   :::|||   |||
Db     562 SerThrLeuSer-----------AlaThrSerThrProHisThrSerThrSer 575
```

Search completed: May 16, 2003, 12:38:15
Job time : 131.5 secs

OM nucleic - protein search, using frame_plus_n2p model

```
Run on:          May 16, 2003, 12:15:31 ; Search time 36.5 Seconds
                                            (without alignments)
                                            4861.253 Million cell updates/sec

Title:           US-10-008-355-1
Perfect score:   3820
Sequence:        1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters:      225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110337_9320/app_query.fasta_1.2311
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_47_@runat_16052003_110337_9320 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       SwissProt_40:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	154.5	4.0	1609	1	FIG2_YEAST	P25653 saccharomyc
	2	145	3.8	1367	1	AMYH_YEAST	P08640 saccharomyc
	3	143.5	3.8	776	1	YLA4_CAEEL	Q05036 caenorhabdi
	4	143.5	3.8	1802	1	HKR1_YEAST	P41809 saccharomyc
	5	139	3.6	4543	1	LRP1_CHICK	P98157 gallus gall
	6	137.5	3.6	1140	1	YM96_YEAST	Q04893 saccharomyc
c	7	132	3.3	571	1	PT1_CHLTR	O84340 chlamydia t
	8	127.5	3.3	725	1	AGA1_YEAST	P32323 saccharomyc
	9	126.5	3.3	3178	1	YS89_CAEEL	Q09624 caenorhabdi
	10	122	3.2	1104	1	COLA_CLOPE	P43153 clostridium
	11	121	3.2	1628	1	NAGH_CLOPE	P26831 clostridium
	12	118	3.1	698	1	EFG_VIBCH	Q9kuz7 vibrio chol
	13	117.5	3.1	1024	1	HLYA_ECOLI	P08715 escherichia
	14	117	3.1	1147	1	TEA1_SCHPO	P87061 schizosacch
	15	116.5	3.0	1169	1	YK82_YEAST	P36170 saccharomyc
c	16	116	2.9	495	1	MURE_XYLFA	Q9pf85 xylella fas
	17	116	3.0	652	1	IRGA_VIBCH	P27772 vibrio chol
	18	116	3.0	952	1	TOP1_STRCO	Q9x909 streptomyce
	19	114	3.0	620	1	DNAK_PORPU	P30723 porphyra pu
	20	114	3.0	734	1	GLGB_AGRTU	P52979 agrobacteri
	21	114	3.0	747	1	VIVC_BPT7	P03725 bacteriopha
	22	113	3.0	507	1	YG46_YEAST	P53301 saccharomyc
	23	113	3.0	1047	1	RIR1_CHLMU	Q9p193 chlamydia m
	24	113	3.0	1161	1	DAN4_YEAST	P47179 saccharomyc
	25	113	3.0	1273	1	WEB1_YEAST	P38968 saccharomyc
	26	113	3.0	4687	1	PLE1_RAT	P30427 rattus norv
	27	112.5	2.9	657	1	CN16_HAEIN	P44764 haemophilus
	28	112	2.9	556	1	EST2_CAEEL	Q07085 caenorhabdi
	29	112	2.9	1013	1	PPOL_HUMAN	P09874 homo sapien
	30	112	2.9	1504	1	DPOZ_YEAST	P14284 saccharomyc
	31	111.5	2.9	878	1	ACON_RICPR	Q9zcf4 rickettsia
	32	111.5	2.9	922	1	YKL6_CAEEL	P42173 caenorhabdi
	33	111.5	2.9	1023	1	HLY1_ECOLI	P09983 escherichia
	34	111.5	2.9	2410	1	MOK1_SCHPO	Q9usk8 schizosacch
	35	110.5	2.9	498	1	FLID_BACSU	P39738 bacillus su
	36	110.5	2.9	1039	1	SYI_METJA	Q58357 methanococc
	37	109.5	2.9	768	1	PARC_NEIGO	P48374 neisseria g
	38	109	2.9	681	1	MP10_HUMAN	O00566 homo sapien
	39	109	2.9	3672	1	LML2_CAEEL	Q21313 caenorhabdi
	40	108.5	2.8	507	1	YY42_CAEEL	Q18416 caenorhabdi
	41	108.5	2.8	1041	1	DD16_HUMAN	O60231 homo sapien
	42	108	2.8	638	1	NEC2_PIG	Q03333 sus scrofa
	43	108	2.8	1341	1	RPA1_METJA	Q58445 methanococc
	44	107.5	2.8	636	1	YNR6_YEAST	P53882 saccharomyc
	45	107.5	2.8	666	1	UVRB_CLOAB	Q971q2 clostridium

ALIGNMENTS

```
RESULT 1
FIG2_YEAST
ID   FIG2_YEAST     STANDARD;      PRT;  1609 AA.
AC   P25653;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Factor induced gene 2.
GN   FIG2 OR YCR089W OR YCR89W OR YCR1102.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92397594; PubMed=1523889;
RA   Wilson C., Grisanti P., Frontali L.;
RT   "The complete sequence of a 6146 bp fragment of Saccharomyces
RT   cerevisiae chromosome III contains two new open reading frames.";
RL   Yeast 8:569-575(1992).
CC   -!- FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC   -!- INDUCTION: BY MATING PHEROMONES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X59720; CAA42254.1; -.
DR   PIR; S19504; S19504.
DR   PIR; S25345; S25345.
DR   SGD; S0000685; FIG2.
SQ   SEQUENCE   1609 AA;  166049 MW;  7D66AD7F85A7B852 CRC64;

Alignment Scores:
Pred. No.:               0.0162       Length:        1609
Score:                   154.50       Matches:       153
Percent Similarity:      35.18%       Conservative:  101
Best Local Similarity:   21.19%       Mismatches:    254
```

```
Query Match:          4.04%     Indels:     214
DB:                   1         Gaps:        28

US-10-008-355-1 (1-2139) x FIG2_YEAST (1-1609)

Qy     74 ACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGC 133
          |||:::   |||     |||                             ||||||
Db    788 ThrGluValCysSerHisSer---------------------------GluCysThrPro 798

Qy    134 TCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATG 193
          :::   :::   |||      :::    :::|||  ||||||                 :::
Db    799 ThrValIleThrSer------ValThrAlaThrSerSerThr---------------Ile 811

Qy    194 CCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCT 253
          |||   ||||||:::                                           |||
Db    812 ProLeuLeuSerThr-------------------------------------------Ser 817

Qy    254 TTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAA------GCACGGTGGATC 307
             |||      :::   ||||||          ||||||        |||
Db    818 SerSerThrValLeuSerSerThrValSerGluGlyAlaLysAsnProAlaAlaSerGlu 837

Qy    308 ACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGG 367
             ||||||         ||||||       :::          :::
Db    838 ValThrIleAsnThrGlnValSerAlaThrSerGluAlaThrSerThr------------ 853

Qy    368 GTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGC 427
                                                                    |||
Db    854 ----------------------------------------------------------Ser 854

Qy    428 TCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAAC 487
          ::::::||||||   |||
Db    855 ThrGlnValSerAlaThr------------------------------------------ 860

Qy    488 TGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCA 547
                          |||   |||   |||   |||   |||   :::   |||
Db    861 ---------------SerAlaThrAlaThrAlaSerGluSerSerThrThrSerGlnVal 875

Qy    548 ACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTC------GTATGGTAT 601
          :::|||   :::   :::   ||||||:::   ::::::   |||
Db    876 SerThrAlaSerGluThrIleSerThrLeuGlyThrGlnAsnPheThrThrThrGlySer 895

Qy    602 TTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGC 661
          ||||||   |||||||||      |||:::|||   |||   ||||||         |||
Db    896 LeuLeuPheProAlaLeu------SerThrGluMetIleAsnThrThrValValSerArg 913

Qy    662 GTCACACGGGCGACTTCA-----------GCGTATTCCGCGTGTATGCCGGTGCCGACA 709
             |||       |||          :::::::|||   |||:::|||
Db    914 LysThrLeuIleIleSerThrGluValCysSerHisSerLysCysValProThrValIle 933

Qy    710 ACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCG 769
          |||            :::       |||   |||   :::      :::|||
Db    934 ThrGluValValThrSerLysGlyThr------ProSerAsnGlyHisSerSerGlnThr 951

Qy    770 TATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTA 829
                :::|||      :::|||               :::|||:::   :::
Db    952 LeuGlnThrGluAlaValGluValThrLeuSerSerHisGlnThrValThrMetSerThr 971

Qy    830 CGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGA--------- 880
             :::   ::::::   :::            :::::::::      :::
Db    972 GluValCysSerAsnSerIleCysThrProThrValIleThrSerValGlnMetArgSer 991

Qy    881 ---------ACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAG 931
                   |||   |||         :::   ||||||   ||||||   :::|||
Db    992 ThrProPheProTyrLeuThrSerSerThrSerSerSerSerLeuAlaSerThrLysLys 1011

Qy    932 CCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTA 991
                                   ||||:::::   ||||||
Db   1012 ---------------------------SerSerLeuGluAlaSer--------------- 1017

Qy    992 ACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTC 1051

Db   1017 ------------------------------------------------------------ 1017

Qy   1052 GTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTG 1111
                                  |||   ||||||   |||      :::   |||
Db   1018 -----------------SerGluMetSerThrPheSerValSerThr---GlnSerLeu 1030

Qy   1112 TCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACC 1171
             :::|||      |||                          |||   |||   |||
Db   1031 ProLeuAlaPheThrCys-----------------------SerGluLysArgSerThr 1042

Qy   1172 GTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCAC 1231
                                       :::   |||   :::|||
Db   1043 ----------------------------ThrSerValSerGlnTrp------------ 1048

Qy   1232 AGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTG 1291
                   ::::::   :::         :::|||      ||||||||   :::
Db   1049 ---------SerAsnThrValLeuThrAsnThrIleMetSerSerSerSerAsnValIle 1065

Qy   1292 ACGACAAGTACAAAG---------ACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCG 1342
          :::|||:::   |||         ||||||||||||   :::
Db   1066 SerThrAsnGluLysProSerSerThrThrSerProTyrAsn----------------- 1079

Qy   1343 CCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGA 1402
                      :::|||   |||||||||:::||||||
Db   1080 ------------PheSerSerGlyTyrSerLeuProSerSerSerThrProSerGlnTyr 1095

Qy   1403 ATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACA 1462
             ||||||      :::      |||      :::::::|||   |||:::|||
Db   1096 SerLeuSerThr------AlaThrThrThrIleAsnGlyIleLysThrValTyrThrThr 1113

Qy   1463 AGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAA 1522
                |||      :::|||      |||         :::|||      :::   :::
Db   1114 ------TrpCysProLeuAlaGluLysSerThrValAlaAlaSerSerGlnSerSerArg 1131

Qy   1523 AGTTTGCCAAGGCTA------TCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAA 1576
          |||:::   |||          |||   |||   |||:::      ||||||::::::   |
Db   1132 SerValAspArgPheValSerSerSerLys--ProSerSerSerLeuSerGlnThrSerI 1151

Qy   1577 TAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGG 1636
          ||                                        |||   :::
Db   1151 le--------------------------------------GlnTyrThrLeuSerThrA 1158

Qy   1637 GCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCC------------------ 1677
                   :::||||||:::   :::|||
Db   1158 laThrThrThr---IleSerGlyLeuLysThrValTyrThrThrTrpCysProLeuThrS 1177

Qy   1678 --------------GGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCT 1723
                        |||         :::   :::         :::|||::::::
Db   1177 erLysSerThrLeuGlyAlaThrThrGlnThrSerSerThrAlaLysValArgIleThr- 1196

Qy   1724 ACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAG 1783
          :::|||   |||                  ::::::   :::|||
Db   1197 SerAlaSerSerAlaThrSerThrSerIleSerLeuSerThrSerThrGluSerGluSer 1216

Qy   1784 GCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATA 1843
          :::      |||         :::
Db   1217 SerSerGlyTyrLeuSerLysGlyValCysSerGlyThrGluCysThrGlnAspValPro 1236

Qy   1844 TCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATA 1903
          :::   :::||||||         |||:::            |||         :::|||
Db   1237 ThrGlnSerSerSerProAlaSerThrLeuAlaTyrSerProSerValSerThrSerSer 1256

Qy   1904 TCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACT-----------------CCG 1945
          |||   |||:::      |||||||||:::      :::|||                |||
Db   1257 SerSerSerPheSerThrThrThrAlaSerThrLeuThrSerThrHisThrSerValPro 1276

Qy   1946 GTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGG 2005
          :::   |||   |||                  |||:::      |||   :::
Db   1277 LeuLeuProSerSerSer---------------SerIleSerAlaSerSerProSerSer 1291

Qy   2006 AAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACA 2065
                :::   ||||||      :::|||      ::::::   :::         |||
Db   1292 ThrSerLeuLeuSerThrSerLeuProSerProAlaPheThrSerSerThrLeuProThr 1311

Qy   2066 TCCGCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
```

```
         :::           ||||||                   |||   |||:::|||||||||||
Db    1312 AlaThrAlaValSerSerSerThrPheIleAlaSerSerLeuProLeuSerSerLysSer 1331

RESULT 2
AMYH_YEAST
ID    AMYH_YEAST     STANDARD;      PRT;  1367 AA.
AC    P08640; P08068;
DT    01-AUG-1988 (Rel. 08, Created)
DT    01-FEB-1995 (Rel. 31, Last sequence update)
DT    16-OCT-2001 (Rel. 40, Last annotation update)
DE    Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE    glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN    STA1 OR STA2 OR MAL5 OR YIR019C.
OS    Saccharomyces cerevisiae (Baker's yeast).
OC    Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC    Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX    NCBI_TaxID=4932;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=S288c / AB972;
RA    Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA    Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA    Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA    Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA    Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA    Walsh S.V., Whitehead S.;
RL    Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN    [2]
RP    SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX    MEDLINE=87194600; PubMed=3106330;
RA    Yamashita I., Nakamura M., Fukui S.;
RT    "Gene fusion is a possible mechanism underlying the evolution of
RT    STA1.";
RL    J. Bacteriol. 169:2142-2149(1987).
RN    [3]
RP    SEQUENCE OF 1-31 FROM N.A.
RC    STRAIN=SPX101-1C;
RX    MEDLINE=89031230; PubMed=3141213;
RA    Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT    "Similar short elements in the 5' regions of the STA2 and SGA genes
RT    from Saccharomyces cerevisiae.";
RL    FEBS Lett. 239:179-184(1988).
CC    -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC        glucose residues successively from non-reducing ends of the chains
CC        with release of beta-D-glucose.
CC    -!- SIMILARITY: TO S.POMBE SPBC215.13.
CC    -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC    the European Bioinformatics Institute.  There are no  restrictions on  its
CC    use  by  non-profit  institutions as long  as its content  is  in  no  way
CC    modified and this statement is not removed.  Usage  by  and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------------
DR    EMBL; Z38061; CAA86176.1; -.
DR    EMBL; M16164; AAA35014.1; -.
DR    EMBL; M16165; AAA35015.1; -.
DR    EMBL; X13857; CAA32069.1; -.
DR    PIR; B26877; B26877.
DR    PIR; A26877; A26877.
DR    PIR; S48478; S48478.
DR    SGD; S0001458; MUC1.
KW    Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW    Signal; Multigene family.
FT    SIGNAL        1     21       POTENTIAL.
FT    CHAIN        22   1367       GLUCOAMYLASE S1/S2.
FT    DOMAIN      210   1367       SER/THR-RICH.
FT    CARBOHYD    817    817       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT    CARBOHYD    874    874       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ    SEQUENCE   1367 AA;  136110 MW;  91C00E2DBD61AA9D CRC64;
```

```
Alignment Scores:
Pred. No.:                0.0657        Length:         1367
Score:                    145.00        Matches:        160
Percent Similarity:       33.69%        Conservative:   95
Best Local Similarity:    21.14%        Mismatches:     354
Query Match:              3.80%         Indels:         148
DB:                       1             Gaps:           25

US-10-008-355-1 (1-2139) x AMYH_YEAST (1-1367)

Qy      65 CCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAA 124
             |||   ||||||   :::         ::::::  :::|||          :::
Db     355 ProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSer 374

Qy     125 TGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCA 184
                       |||:::                  :::|||   |||:::|||   |||
Db     375 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 394

Qy     185 TTGCCAATGCCGTGGTTATCTTCGGTG------GCGGATGTACCGGTATCACAGTGTCCG 238
             :::|||   |||      |||:::            :::      ||||||:::
Db     395 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThr 414

Qy     239 ATCAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCA 298
                 :::   |||   |||   |||::::::   |||         |||   |||
Db     415 ThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProVal 434

Qy     299 CGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTC 358
                       ||||||      :::               :::            |||
Db     435 ThrSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSer 454

Qy     359 CGATTCCGGGTC---TTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGG 415
                       |||      |||   |||            |||
Db     455 SerAlaProValProThrProSerSerSerThrThrGluSer------------------ 468

Qy     416 TAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGG 475
                 :::               |||   |||        |||      |||   :::
Db     469 ---SerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProValProThr 487

Qy     476 TATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGC 535
              :::   :::         :::                     |||:::|||   :::   |||
Db     488 ProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSer 507

Qy     536 CTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTA 595
              |||   |||               |||||||||:::|||         |||
Db     508 SerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAla------ 525

Qy     596 TGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGT 655
                             ||||||         ||||||            ::::::
Db     526 --------------ProAlaProThrProSerSerSerThrThrGluSerSerSerAla 540

Qy     656 GGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGC 715
                       :::      |||   :::   ||||||      |||||||||||
Db     541 ProValThrSerSerThrThrGluSerSerSerAla------ProValProThrProSer 558

Qy     716 CGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACT-----------TCG 763
                       |||      :::      |||::::::      |||            |||
Db     559 SerSerThrThrGluSerSerSerThrProValThrSerSerThrThrGluSerSerSer 578

Qy     764 CTGCCGTATCCA---------TGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCG 814
              |||   |||                        |||      ::::::   |||   |||:::
Db     579 AlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProValProThr 598

Qy     815 GTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAA 874
              |||   :::            ::::::|||                        |||
Db     599 ProSerSerSerThrThrGluSerSerSerAlaProAlaProThrProSerSerSerThr 618

Qy     875 ACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCA 934
             |||   :::                  :::   ::::::::::::|||
Db     619 ThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSer--------- 635

Qy     935 TGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACT 994
                |||   :::      |||   |||:::                              |||
Db     636 ---AlaProValProThrProSerSerSer------------------------ThrThr 646
```

```
Qy    995 ATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTA 1054
                            |||   |||      |||
Db    647 GluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAla 666

Qy   1055 AGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCC-----------GTAAGAACGGCA 1102
             ||||||            |||   |||   |||            |||   ::::::
Db    667 ProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSer 686

Qy   1103 AGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCA 1162
                   |||   :::                 |||                     |||
Db    687 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 706

Qy   1163 AGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTC 1222
             ||||||                            |||:::   ||||||
Db    707 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValPro-------- 723

Qy   1223 GTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCA 1282
                   |||      |||                    :::      |||   ||||||
Db    724 ThrProSerSerSerThrThrGluSerSerSerAlaProValProThrProSerSerSer 743

Qy   1283 AATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCG 1342
                   ::::::|||         ||||||            :::               |||
Db    744 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 763

Qy   1343 ---------CCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATA 1393
                   |||   |||                  |||   |||      :::|||
Db    764 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValProThrProSer 783

Qy   1394 TATTCAAGAATGTAATCG------------------ACAAGAAATTCAAAGGCGACACGA 1435
             |||         |||                  |||   :::|||      |||
Db    784 SerSerThrThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGlu 803

Qy   1436 AGAAGTATG-----------CAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCG 1483
             |||:::                 |||   |||:::   :::
Db    804 SerSerValAlaProValProThrProSerSerSerSerAsnIle--------------- 818

Qy   1484 ACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGA 1543
          |||||||||   |||   |||:::|||::::::      :::
Db    819 ThrSerSerAlaProSerSerThrProPheSerSerSerThr----------------- 832

Qy   1544 AAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCG 1603

Db    833 -------------------------------------------------------GluSer 834

Qy   1604 ATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGC 1663
                   |||:::|||   |||                  :::      |||
Db    835 SerSerValProValProThrProSerSerSerThrThrGluSerSerSerAlaProVal 854

Qy   1664 GTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACT---------TCACCATGC 1714
                   |||   :::                 :::||||||         |||
Db    855 SerSerSerThrThrGluSerSerValAlaProValProThrProSerSerSerSerAsn 874

Qy   1715 GTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCT----------- 1762
          :::   ::::::|||||||||                  ::::::|||
Db    875 IleThrSerSerAlaProSer-----------------SerIleProPheSerSerThr 888

Qy   1763 -----------------GGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGC 1804
                            |||||||||:::         :::      |||      |||
Db    889 ThrGluSerPheSerThrGlyThrThrValThrProSerSerSerLysTyrProGlySer 908

Qy   1805 AGGATCCTAAGAGCGATGAGT--------------------TTGCCGTACAGGAGAATA 1843
          :::         ::::::|||                     :::|||   :::
Db    909 GlnThrGluThrSerValSerSerThrThrGluThrThrIleValProThrLysThrThr 928

Qy   1844 TCCTCGACCTCTTCCGCACCAAAAACTATGGTC-------------------------- 1876
          :::|||   ::::::         |||   :::
Db    929 ThrSerValThrThrProSerThrThrThrIleThrThrThrValCysSerThrGlyThr 948

Qy   1877 --------------------------GCTATGCCGAGAACGGTCAGCTCCATATCGCTT 1909
                                           |||:::||||||::::::
Db    949 AsnSerAlaGlyGluThrThrSerGlyCysSerProLysThrValThrThr--------- 965

Qy   1910 TCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACG 1969
          :::      |||||||||:::   :::||||||      :::      :::|||   |||
Db    966 ThrValProThrThrThrThrThrSerValThrThrSerSerThrThrThrIleThrThr 985

Qy   1970 GCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGT 2029
             |||                |||      |||                  ||||||
Db    986 ThrVal------------CysSerThrGlyThrAsnSerAlaGlyGluThrThrSerGly 1001

Qy   2030 TCGAACCCG--------------------ATCTGCAGCGCACAATCAGCGTGGACATCC 2068
             :::|||                         ||||||                  |||
Db   1002 CysSerProLysThrIleThrThrThrValProCysSerThrSer---------ProSer 1018

Qy   2069 GCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCC 2119
             |||   |||         |||         :::   ||||||::::::
Db   1019 GluThrAlaSerGluSerThrThrThrSerProThrThrProValThrThr 1035
```

RESULT 3
YLA4_CAEEL

```
ID   YLA4_CAEEL     STANDARD;      PRT;   776 AA.
AC   Q05036;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 86.9 kDa protein C30C11.4 in chromosome III.
GN   C30C11.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY. STRONG,
CC       TO YEAST MSI3.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L09634; AAA27967.1; -.
DR   PIR; S44784; S44784.
DR   WormPep; C30C11.4; CE00103.
DR   InterPro; IPR001023; Hsp70.
DR   Pfam; PF00012; HSP70; 1.
DR   ProDom; PD000089; Hsp70; 1.
DR   PROSITE; PS00297; HSP70_1; FALSE_NEG.
DR   PROSITE; PS00329; HSP70_2; 1.
DR   PROSITE; PS01036; HSP70_3; 1.
KW   Hypothetical protein; ATP-binding; Multigene family.
SQ   SEQUENCE   776 AA;  86896 MW;  20FA975FB649FA9C CRC64;
```

```
Alignment Scores:
Pred. No.:          0.0687        Length:        776
Score:              143.50        Matches:       118
```

```
Percent Similarity:      37.48%       Conservative:   87
Best Local Similarity:   21.57%       Mismatches:     189
Query Match:             3.76%        Indels:         153
DB:                      1            Gaps:           27

US-10-008-355-1 (1-2139) x YLA4_CAEEL (1-776)

Qy     113 ATCTGGATCGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCG 172
           ||||||   :::             ||||||          |||
Db     236 IleTrpPheAsp-----------AlaLeuIleArgGluHisPheArgLysGluPheLys 251

Qy     173 ACAAGCCGTCCATTG--------------CCAATGCCGTGGTTATCTTCGGTGGCGGAT 217
           |||        :::               |||  |||||||||     :::    :::
Db     252 ThrLysTyrGlyIleAspAlaAlaThrSerProArgProTrpLeuArgLeuLeuAspGlu 271

Qy     218 GTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCACTGCGGATACG 277
                          :::      :::   |||       |||   ||||||  |||
Db     272 CysGluArgValLysLysGlnMetSerAlaAsnGlnThrProIle-ProLeuAsnIleGl 291

Qy     278 GTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTC 337
            |||:::   |||  |||               ::::::    |||      |||
Db     291 uCysPheMetGluAspLysAspValThrGlyLysMetGln-ArgGlnGluPhe------- 308

Qy     338 GCACGATGGGTGAGGAGCTTCCGATTCCG------------------------------ 366
                     |||:::|||     |||
Db     309 ----------GluAspLeuAlaAlaProIlePheAsnArgIleLysGlnValLeuIleA 325

Qy     367 -------------GGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACA 412
                        |||:::|||:::|||                :::      |||:
Db     325 snLeuPheAlaAspGlyValSerIleLys--------------ProGluGluIleAspG 340

Qy     413 AGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGG 472
           :::::|||               |||                |||:::         :::|
Db     340 luIleGlu---------------IleValGlyGlySerSerArgIleProMetIleArgG 355

Qy     473 AGGTATGCCAAGAACTGGCCAAAAAAGAA---------------AATGCAGACGAG---- 513
           |||:::   ::::::|||      ||||||                |||   ||||||
Db     355 luIleValLysAspLeuPheGlyLysGluProLysThrThrMetAsnGlnAspGluAlaV 375

Qy     514 -----------------AACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAAT 556
                            |||     |||:::   |||    :::          |||:
Db     375 alAlaArgGlyAlaAlaMetGlnCysAlaIleLeuSerProThrPheArgValArgGluP 395

Qy     557 ACTTCCTCATCGTCTACGATGTATTCAAGGAC---------------GTTCGTATGGTAT 601
           ::                            ||||||           :::|||:::   :
Db     395 he------------------AlaIleLysAspThrGlnProTyrArgIleArgLeuSerT 409

Qy     602 TTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGC 661
           ::             :::|||  |||:::   ||||||:::   |||   :::      ||||
Db     409 rp---------AsnSerThrGlyGluAsnGlyGlyGluAsnAspValPheSer---ProA 425

Qy     662 GTCACACGGGCGACTTCAGC------GTATTCCGCGTGTATGCCGGTGCCGACAACCGGC 715
           ||         ||||||     |||       :::    :::|||      |||
Db     425 rgAspGluValProPheSerLysLeuValSerLeuLeuArgSerGlyProPheAsnValG 445

Qy     716 CGGCCGAATACAGCAAGGACAAT---AAACCCTATAAGCCCGTTTACTTCGCTGCCGTAT 772
              |||   |||::::::   |||         |||:::        |||:::       :::
Db     445 luAlaHisTyrAlaGlnProAsnValValProHisAsnGlnValHisIleGlySerTrpL 465

Qy     773 CCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGG 832
              :::    |||    :::                              ||||||:::
Db     465 ysValAsnGlyAlaArg----------------------------ProGlyAlaAsp- 474

Qy     833 ATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTA 892
                               |||           ::::::                  ::::
Db     475 -------------------GlyGlyAsnGlnLysVal------------------LysV 482

Qy     893 TCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTA 952
            ::::::||||||                |||||||:::      |||
Db     482 alLysValArgValAsnProAspGlyIlePheThrIleAla------------------- 495

Qy     953 CCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCG 1012
                                          ||||||       ::::::        |||
Db     496 ------------------------------SerAlaThrMetTyrGluProArgIleV 505

Qy    1013 GTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAG 1072
                                   ::::::|||    |||      :::    |||      |
Db     505 alGluGluValProAlaGluAlaMetGluValAspGlyAspAlaLysThrGluAlaProA 525

Qy    1073 CATTC---GCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGT 1129
           ||          :::   :::::::|||                 |||      :::|||:::
Db     525 laGluProLeuGluProValLysLysThrLysLeuValProValAspLeuGluValIleG 545

Qy    1130 CTTCTCTCGAAAAGGCTTAT---------------------------------------- 1149
            |||:::       :::|||
Db     545 luSerIleProValSerTyrAspValGlnLysPheHisAsnLeuGluLeuGlnMetGlnG 565

Qy    1150 -----------AAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGC 1198
                      :::|||    |||||||||:::   :::    :::    |||   |||   :
Db     565 luSerAspAlaArgGluLysAlaLysAlaAspAlaLysAsnSer---LeuGluGluTyrV 584

Qy    1199 TCTTCGGTGGTACCGAGGTGGTT------CGTTTTGCACAGTTTGCCAACGCATTGGCTA 1252
           :::::         :::   |||      ::::::|||:::|||              |||
Db     584 alTyrGluMetArgAspLysValSerAspGlnTyrAlaGluPheIleThrProAlaAlaA 604

Qy    1253 CAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACC 1312
                   |||            :::|||   |||    :::|||            |||
Db     604 la------AspGluPheArgSerValLeuThrSerThrGluAspTrpLeuTyrAspGluG 622

Qy    1313 TCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCC 1372
                 :::|||                      |||:::          :::|||:::
Db     622 lyGluAspAlaGluArg--------------------AspValTyrGluLysArgLeu- 634

Qy    1373 CTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACA 1432
                :::|||        :::             |||:::::::::       :::    :::|
Db     635 -----SerGluLeuLysAlaValGlyThrProValValGluArgTyrArgGluSerGluT 653

Qy    1433 CGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTT-----------CCTTATA 1480
            |||:::|||    |||             ||||||:::|||:::::::          |||
Db     653 hrArgLysProAlaPheAspSerPheAspGlnSerIleMetArgValArgLysAlaTyrG 673

Qy    1481 GCGAC--------------AAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGT 1525
              |||                :::|||   |||   |||    :::   ||||||
Db     673 luAspTyrAlaAsnGlyGlyProThrTyrAlaHisLeuAspSerLysGluMetGluLysV 693

Qy    1526 TTGCCAAGGCTATCGAG 1542
                ||||||||||
Db     693 alIleAsnAlaIleGlu 698

RESULT 4
HKR1_YEAST
ID   HKR1_YEAST     STANDARD;      PRT;  1802 AA.
AC   P41809;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hansenula MRAKII killer toxin-resistant protein 1 precursor.
GN   HKR1 OR YDR420W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YNN 295;
RX   MEDLINE=94156857; PubMed=8113191;
RA   Kasahara S., Yamada H., Mio T., Shiratori Y., Miyamoto C.,
RA   Yabe T., Nakajima T., Ichishima E., Furuichi Y.;
RT   "Cloning of the Saccharomyces cerevisiae gene whose overexpression
RT   overcomes the effects of HM-1 killer toxin, which inhibits
RT   beta-glucan synthesis.";
RL   J. Bacteriol. 176:1488-1499(1994).
CC   -!- FUNCTION: COULD REGULATE BETA-GLUCAN SYNTHESIS. OVEREXPRESSION
CC       PROVIDES RESISTANCE TO HM-1 KILLER TOXIN.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC   -!- PTM: COULD BE O-GLYCOSYLATED IN SERINE/THREONINE RICH DOMAIN.
```

```
CC   -!- SIMILARITY: SOME, TO YEAST MSB2.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; S69101; AAB30051.1; -.
DR   SGD; S0002828; HKR1.
KW   Glycoprotein; Transmembrane; Repeat; Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1802       HANSENULA MRAKII KILLER TOXIN-RESISTANT
FT                                PROTEIN 1.
FT   TRANSMEM   1486   1506       POTENTIAL.
FT   DOMAIN       23   1478       SER/THR-RICH.
FT   DOMAIN      453    788       12 X 28 AA TANDEM REPEATS OF S-[AV]-[P]-
FT                                V-A-V-S-S-T-Y-T-S-S-P-S-A-P-A-A-I-S-S-T-
FT                                Y-T-S-S-P.
FT   REPEAT      453    480       1 (APPROXIMATE).
FT   REPEAT      481    508       2.
FT   REPEAT      509    536       3.
FT   REPEAT      537    564       4.
FT   REPEAT      565    592       5.
FT   REPEAT      593    620       6.
FT   REPEAT      621    648       7.
FT   REPEAT      649    676       8.
FT   REPEAT      677    704       9.
FT   REPEAT      705    732       10.
FT   REPEAT      733    760       11.
FT   REPEAT      761    788       12.
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1252   1252       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1293   1293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1342   1342       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1400   1400       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1802 AA;  188890 MW;  E344CA6469785A24 CRC64;

Alignment Scores:
Pred. No.:              0.0905        Length:          1802
Score:                  143.50        Matches:         167
Percent Similarity:     36.03%        Conservative:    105
Best Local Similarity:  22.12%        Mismatches:      282
Query Match:            3.76%         Indels:          202
DB:                     1             Gaps:            33

US-10-008-355-1 (1-2139) x HKR1_YEAST (1-1802)

Qy       1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
             :::   :::|||:::|||   ||||||||      :::
Db       2 ValSerLeuLysIleLysLysIleLeuLeuLeuValSer-------------------- 14

Qy      61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
                                     |||||||||   :::        :::    |||
Db      15 -------------------------LeuLeuAsnAlaIleGluAlaTyrSerAsnAsp 25

Qy     121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
              :::             :::       ::::::|||        |||:::
Db      26 ThrIleTyrSerThrSerTyrAsnAsnGlyIleGluSerThrProSerTyrSerThrSer 45

Qy     181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
             :::|||::::::                          ||||||    :::     :::
Db      46 AlaIleSerSer------------------------ThrGlySerSerAsnLysGlu 56

Qy     241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGC--- 297
                                                   ||||||   |||    |||
Db      57 ---------------------------------AsnAlaIleThrSerSerSerGlu 64

Qy     298 ------ACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAG 351
                 |||:::          |||         |||      |||    |||:::    :::
Db      65 ThrThrThrMetAlaGlyGlnTyrGlyGluSerGly------SerThrThrIleMetAsp 82

Qy     352 GAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGAC 411
           |||         |||   |||   :::|||                 :::   ||||||
Db      83 GluGlnGlu---ThrGlyThrSerSerGlnTyr-----------IleSerValThrThr 97

Qy     412 AAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAG 471
                 :::                   :::|||   |||      ::::::|||:::
Db      98 ThrThrGln--------------ThrSerAspThrMetSerSerValLysLysSerThr 112

Qy     472 GAGGTATGCCAAGAACTGGCCAAA-----------------AAAGAAAATGCAGACGAG 513
           |||:::          :::                       :::      :::||||||
Db     113 GluIleAlaThrProSerSerSerIleValProThrProLeuGlnSerTyrSerAspGlu 132

Qy     514 AACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTAC 573
           :::|||        :::              |||:::|||
Db     133 SerGln---------IleSerGlnThrLeuSerHisAsn-------------------- 142

Qy     574 GATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGA 633
                                                 |||   ||||||    :::
Db     143 ------------------------------------ProLysSerValAlaGluSerAsp 150

Qy     634 GGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTG 693
              |||||| |||                          :::|||      |||:::
Db     151 SerAspThr-Thr--------------------------SerSerGluSerSerSerSe 161

Qy     694 TATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCC 753
             :::   :::   |||:::                 |||   |||:::||||||
Db     161 rValIleIleSerThrSerAspSerSerAlaValProArgGluIleSerProIleIleTh 181

Qy     754 GTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATC 813
                  |||         :::   |||           |||              ||
Db     181 rThrAspSerGlnIleSerLysGluGluGlyThrLeuAlaGlnThr---------SerSe 198

Qy     814 GGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAA 873
           |:::|||         |||||||||      :::        |||   :::|||   |||
Db     198 rIleSerGluThrThrArgIleAlaGlnMetValThrArgValSerGlnIleSerSerIl 218

Qy     874 AACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATC----TGGAAGGA 929
             |||         :::      |||:::      |||           :::
Db     218 eThrAlaAlaSerThrIleAspGlyPheSerSerGluSerThrGlnThrAspPheSerAs 238

Qy     930 AGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGC 989
                 :::|||   :::   :::        :::||||||        :::|||   ::
Db     238 nThrValSerPheGluAsnSerValGluGluGluTyrAlaMetSerLysSerGlnLeuSe 258

Qy     990 TAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGG 1049
           :        :::   :::|||      :::        |||                ||
Db     258 rGluSerTyrSerSerSerSerThrValTyrSerGly--------------------Gl 271

Qy    1050 TCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGC 1109
           |         |||:::                              |||       |||
Db     271 yGluSerThrAlaAsp-------------------------LysThrSerSerSerPr 282

Qy    1110 TGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAA 1169
             :::               |||       :::              |||   ::::::::::
Db     282 oIleThrSerPheSerSerSerTyrSerGlnThrThrSerThrGluThrSerGluSerSe 302

Qy    1170 CCGTGAGATGACTTATTTGAGC------GAGACGCTCTTCGGTGGTACCGAGGTGGTTCG 1223
            :|||            :::|||          ::::::         |||    :::
Db     302 rArgValAlaValGlyValSerArgProSerSerIleThrGlnThrThrSerIleAspSe 322

Qy    1224 TTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGT-------- 1275
            |||:::                |||:::|||      |||      ||||||
Db     322 rPheSerMetSerGluValGluLeuSerThrTyrTyrAspLeuSerAlaGlyAsnTyrPr 342

Qy    1276 ----------------ATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTC 1319
                              :::         |||:::||||||      ||||||
Db     342 oAspGlnGluLeuIleVal-AspArgProAlaThrSerSerThrAlaGluThrSerSerG 362

Qy    1320 GCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGA 1379
              ::::::                                  |||||||||      :
Db     362 luAlaSer--------------------------------GlnGlyValSerArgGluS 371
```

```
Qy   1380 CAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAAT---------TCAAAGGCGA 1430
          :::::::        |||   :::  ||||||  |||        |||  |||:
Db    371 erAsnThrPheAlaValSerSerIle---SerThrThrAsnPheIleValSerSerAlaS 390

Qy   1431 CACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTG--------------------T 1469
          ::         :::  ||||||   :::      |||
Db    390 erAspThrValValSerThrSerSerThrAsnThrValProTyrSerSerValHisSerT 410

Qy   1470 GGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTG- 1528
             |||:::  ||||||||             ||||||   :::       ||||||
Db    410 hrPheValHisAlaThrSer-----------SerSerThrTyrIleSerSerSerLeuT 426

Qy   1529 --------CCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGC 1580
                  |||  ||||||      :::         |||      |||
Db    426 yrSerSerProSerLeuSerAlaSerValSerSer---HisPheGlyValAla------- 442

Qy   1581 TGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAA 1640
                                                 |||   |||          |
Db    443 ------------------------------------ProPheProSerAlaTyrIleS 450

Qy   1641 GCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGC 1700
           ||      |||:::||||||   |||     |||
Db    450 erPheSerSerValProValAlaValSerSerThr--------------------TyrT 463

Qy   1701 CAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGC 1760
             :::||||||                ||||||         :::       :::||||
Db    463 hrSerSerProSerAlaSerValValValProSerAlaTyrAlaSerSerProSerValP 483

Qy   1761 CTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGA 1820
           ||           :::       :::      |||      |||
Db    483 ro-----------ValAlaValSerSerThrTyrThrSerSer--------------P 494

Qy   1821 TGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTA 1880
             |||    |||              |||:::    :::|||:::|||        ||||||:
Db    494 roSerAlaProAlaAlaIleSerSerThrTyrThrSerSerProSerAlaProValAlaV 514

Qy   1881 TGCCGAGAACGGTCAGCTCC--------------------ATATCGCTTTCCTATCGAA 1919
           ::        |||    :::|||                    ||||||   :::|||
Db    514 alSerSerThrTyrThrSerSerProSerAlaProAlaAlaIleSerSerThrTyr---- 532

Qy   1920 CAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGAT 1979
                 ||||||            |||||||||      |||       |||:::        |
Db    533 -----ThrSerSerProSerAlaProValAlaValSerSerThrTyrThrSerSerProS 551

Qy   1980 CGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGA 2039
           ||          ::::::::::::               |||||||||    :::
Db    551 erAlaProAlaAlaIleSerSer------------ThrTyrThrSerSerProSerAlaP 567

Qy   2040 TCTGCAGCGCACAATCAGCGTGGACATCC--------------GCTACGTTCTCTTCAT 2084
                :::        |||    :::||||||                 |||      ||||||
Db    567 roValAlaValSerSerThrTyrThrSerSerProSerAlaProAlaAlaIleSerSerT 587

Qy   2085 GATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
                |||:::       |||||||||||||:::        |||
Db    587 hrTyrThrSerSerProSerAlaProValAlaValSerSer 600

RESULT 5
LRP1_CHICK
ID   LRP1_CHICK     STANDARD;      PRT;  4543 AA.
AC   P98157;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Low-density lipoprotein receptor-related protein 1 precursor (LRP)
DE   (Alpha-2-macroglobulin receptor) (A2MR).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn; TISSUE=Liver, and Ovary;
RX   MEDLINE=94103212; PubMed=7506255;
RA   Nimpf J., Stifani S., Bilous P.T., Schneider W.J.;
RT   "The somatic cell-specific low density lipoprotein receptor-related
RT   protein of the chicken. Close kinship to mammalian low density
RT   lipoprotein receptor gene family members.";
RL   J. Biol. Chem. 269:212-219(1994).
CC   -!- FUNCTION: INVOLVED IN THE PLASMA CLEARANCE OF CHYLOMICRON REMNANTS
CC       AND ACTIVATED ALPHA 2-MACROGLOBULIN, AS WELL AS THE LOCAL
CC       METABOLISM OF COMPLEXES BETWEEN PLASMINOGEN ACTIVATORS AND THEIR
CC       ENDOGENOUS INHIBITORS. BINDS VITELLOGENIN, CALCIUM AND ALPHA 2-
CC       MACROGLOBULIN.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: SOMATIC.
CC   -!- PTM: CLEAVED INTO A 85 kDa MEMBRANE-SPANNING SUBUNIT (LRP-85) AND
CC       A 515 kDa LARGE EXTRACELLULAR DOMAIN (LRP-515) THAT REMAINS NON-
CC       COVALENTLY ASSOCIATED.
CC   -!- SIMILARITY: CONTAINS 22 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 31 LDL-RECEPTOR CLASS A DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X74904; CAA52870.1; -.
DR   HSSP; Q07954; 1CR8.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR002172; LDL_recept_A.
DR   InterPro; IPR000033; Ldl_receptor_rep.
DR   Pfam; PF00008; EGF; 14.
DR   Pfam; PF00057; ldl_recept_a; 31.
DR   Pfam; PF00058; ldl_recept_b; 33.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00179; EGF_CA; 3.
DR   SMART; SM00001; EGF_like; 18.
DR   SMART; SM00192; LDLa; 31.
DR   SMART; SM00135; LY; 34.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS01209; LDLRA_1; 27.
DR   PROSITE; PS50068; LDLRA_2; 31.
KW   Receptor; Transmembrane; Repeat; Endocytosis; Glycoprotein;
KW   Signal; Calcium-binding; EGF-like domain; Coated pits;
KW   Alternative splicing.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   4543       LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED
FT                                PROTEIN 1.
FT   DOMAIN       22   4419       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   4420   4443       POTENTIAL.
FT   DOMAIN     4444   4543       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       27     68       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN       72    112       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      113    151       EGF-LIKE 1.
FT   DOMAIN      152    191       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      476    522       EGF-LIKE 3.
FT   DOMAIN      801    841       EGF-LIKE 4.
FT   DOMAIN      850    890       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      891    931       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      932    971       LDL-RECEPTOR CLASS A 5.
FT   DOMAIN      972   1011       LDL-RECEPTOR CLASS A 6.
FT   DOMAIN     1011   1051       LDL-RECEPTOR CLASS A 7.
FT   DOMAIN     1058   1097       LDL-RECEPTOR CLASS A 8.
```

```
FT   DOMAIN     1100   1140       LDL-RECEPTOR CLASS A 9.
FT   DOMAIN     1141   1180       LDL-RECEPTOR CLASS A 10.
FT   DOMAIN     1181   1220       EGF-LIKE 5.
FT   DOMAIN     1221   1260       EGF-LIKE 6.
FT   DOMAIN     1534   1577       EGF-LIKE 7.
FT   DOMAIN     1842   1883       EGF-LIKE 8.
FT   DOMAIN     2151   2191       EGF-LIKE 9.
FT   DOMAIN     2472   2512       EGF-LIKE 10.
FT   DOMAIN     2516   2557       LDL-RECEPTOR CLASS A 11.
FT   DOMAIN     2558   2596       LDL-RECEPTOR CLASS A 12.
FT   DOMAIN     2597   2635       LDL-RECEPTOR CLASS A 13.
FT   DOMAIN     2636   2684       LDL-RECEPTOR CLASS A 14.
FT   DOMAIN     2688   2730       LDL-RECEPTOR CLASS A 15.
FT   DOMAIN     2730   2769       LDL-RECEPTOR CLASS A 16.
FT   DOMAIN     2770   2812       LDL-RECEPTOR CLASS A 17.
FT   DOMAIN     2814   2853       LDL-RECEPTOR CLASS A 18.
FT   DOMAIN     2854   2897       LDL-RECEPTOR CLASS A 19.
FT   DOMAIN     2900   2938       LDL-RECEPTOR CLASS A 20.
FT   DOMAIN     2939   2978       EGF-LIKE 11.
FT   DOMAIN     2979   3019       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     3287   3328       EGF-LIKE 13.
FT   DOMAIN     3329   3368       LDL-RECEPTOR CLASS A 21.
FT   DOMAIN     3369   3407       LDL-RECEPTOR CLASS A 22.
FT   DOMAIN     3408   3447       LDL-RECEPTOR CLASS A 23.
FT   DOMAIN     3448   3488       LDL-RECEPTOR CLASS A 24.
FT   DOMAIN     3489   3530       LDL-RECEPTOR CLASS A 25.
FT   DOMAIN     3531   3569       LDL-RECEPTOR CLASS A 26.
FT   DOMAIN     3570   3608       LDL-RECEPTOR CLASS A 27.
FT   DOMAIN     3608   3646       LDL-RECEPTOR CLASS A 28.
FT   DOMAIN     3649   3689       LDL-RECEPTOR CLASS A 29.
FT   DOMAIN     3690   3730       LDL-RECEPTOR CLASS A 30.
FT   DOMAIN     3736   3776       LDL-RECEPTOR CLASS A 31.
FT   DOMAIN     3779   3821       EGF-LIKE 14.
FT   DOMAIN     3822   3859       EGF-LIKE 15.
FT   DOMAIN     4146   4182       EGF-LIKE 16.
FT   DOMAIN     4195   4231       EGF-LIKE 17.
FT   DOMAIN     4231   4267       EGF-LIKE 18.
FT   DOMAIN     4267   4303       EGF-LIKE 19.
FT   DOMAIN     4303   4339       EGF-LIKE 20.
FT   DOMAIN     4339   4374       EGF-LIKE 21.
FT   DOMAIN     4372   4409       EGF-LIKE 22.
FT   SITE       3939   3942       RECOGNITION SITE FOR PROTEOLYTICAL
FT                                PROCESSING (POTENTIAL).
FT   SITE       4501   4506       ENDOCYTOSIS SIGNAL (POTENTIAL).
FT   DISULFID     29     42       BY SIMILARITY.
FT   DISULFID     36     55       BY SIMILARITY.
FT   DISULFID     49     66       BY SIMILARITY.
FT   DISULFID     74     87       BY SIMILARITY.
FT   DISULFID     81    100       BY SIMILARITY.
FT   DISULFID     94    110       BY SIMILARITY.
FT   DISULFID    117    126       BY SIMILARITY.
FT   DISULFID    122    135       BY SIMILARITY.
FT   DISULFID    137    150       BY SIMILARITY.
FT   DISULFID    156    166       BY SIMILARITY.
FT   DISULFID    162    175       BY SIMILARITY.
FT   DISULFID    177    190       BY SIMILARITY.
FT   DISULFID    480    495       BY SIMILARITY.
FT   DISULFID    491    506       BY SIMILARITY.
FT   DISULFID    508    521       BY SIMILARITY.
FT   DISULFID    805    816       BY SIMILARITY.
FT   DISULFID    812    825       BY SIMILARITY.
FT   DISULFID    827    840       BY SIMILARITY.
FT   DISULFID    852    864       BY SIMILARITY.
FT   DISULFID    859    877       BY SIMILARITY.
FT   DISULFID    871    888       BY SIMILARITY.
FT   DISULFID    893    905       BY SIMILARITY.
FT   DISULFID    900    918       BY SIMILARITY.
FT   DISULFID    912    929       BY SIMILARITY.
FT   DISULFID    934    946       BY SIMILARITY.
FT   DISULFID    941    959       BY SIMILARITY.
FT   DISULFID    953    969       BY SIMILARITY.
FT   DISULFID    974    987       BY SIMILARITY.
FT   DISULFID    982   1000       BY SIMILARITY.
FT   DISULFID    994   1009       BY SIMILARITY.
FT   DISULFID   1013   1025       BY SIMILARITY.
FT   DISULFID   1020   1038       BY SIMILARITY.
FT   DISULFID   1032   1049       BY SIMILARITY.
FT   DISULFID   1060   1073       BY SIMILARITY.
FT   DISULFID   1067   1086       BY SIMILARITY.
FT   DISULFID   1080   1095       BY SIMILARITY.
FT   DISULFID   1102   1116       BY SIMILARITY.
FT   DISULFID   1110   1129       BY SIMILARITY.
FT   DISULFID   1123   1138       BY SIMILARITY.
FT   DISULFID   1143   1157       BY SIMILARITY.
FT   DISULFID   1150   1170       BY SIMILARITY.
FT   DISULFID   1164   1180       BY SIMILARITY.
FT   DISULFID   1183   1194       BY SIMILARITY.
FT   DISULFID   1190   1204       BY SIMILARITY.
FT   DISULFID   1206   1219       BY SIMILARITY.
FT   DISULFID   1225   1235       BY SIMILARITY.
FT   DISULFID   1231   1244       BY SIMILARITY.
FT   DISULFID   1246   1259       BY SIMILARITY.
FT   DISULFID   1538   1551       BY SIMILARITY.
FT   DISULFID   1547   1561       BY SIMILARITY.
FT   DISULFID   1563   1576       BY SIMILARITY.
FT   DISULFID   1846   1857       BY SIMILARITY.
FT   DISULFID   1853   1867       BY SIMILARITY.
FT   DISULFID   1869   1882       BY SIMILARITY.
FT   DISULFID   2155   2166       BY SIMILARITY.
FT   DISULFID   2162   2176       BY SIMILARITY.
FT   DISULFID   2178   2190       BY SIMILARITY.
FT   DISULFID   2476   2487       BY SIMILARITY.
FT   DISULFID   2483   2497       BY SIMILARITY.
FT   DISULFID   2499   2511       BY SIMILARITY.
FT   DISULFID   2518   2531       BY SIMILARITY.
FT   DISULFID   2526   2544       BY SIMILARITY.
FT   DISULFID   2538   2555       BY SIMILARITY.
FT   DISULFID   2560   2572       BY SIMILARITY.
FT   DISULFID   2567   2585       BY SIMILARITY.
FT   DISULFID   2579   2594       BY SIMILARITY.
FT   DISULFID   2599   2611       BY SIMILARITY.
FT   DISULFID   2606   2624       BY SIMILARITY.

Alignment Scores:
Pred. No.:              0.244       Length:         4543
Score:                  139.00      Matches:        139
Percent Similarity:     31.52%      Conservative:   69
Best Local Similarity:  21.06%      Mismatches:     196
Query Match:            3.64%       Indels:         256
DB:                     1           Gaps:           37

US-10-008-355-1 (1-2139) x LRP1_CHICK (1-4543)

Qy    268 TGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGT 327
          ||||||  |||:::   :::          :::  :::|||          |||
Db   3765 CysGlyAspGlySerAspGluLysSerCysSerHisGluHisLysSerTyr---AspCys 3783

Qy    328 TTCGTTTCTCGCACGATG---GGTGAGGAGCTTCCG---ATTCCGGGTCTTTCCGTGAAG 381
             :::  ||||||  |||:::|||         |||         |||
Db   3784 MetThrAsnThrThrMetCysGlyAspGluAlaGlnCysIleGlnAlaGlnSerSerThr 3803

Qy    382 TATCTG---------CGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAG 432
          |||          |||     ||||||  ||||||       |||
Db   3804 TyrCysThrCysArgArgGlyPheGlnLysValProAspLysAsnSerCysGln------ 3821

Qy    433 GGTATCACTGACGAGATGGAGCGTCTGCGC---------------------------- 462
                   |||      |||  ||||||
Db   3822 ---------AspValAsnGluCysLeuArgPheGlyThrCysSerGlnLeuCysAsnAsn 3838

Qy    463 ---AAAGCTCAGGAGGTATGC---------CAAGAACTGGCCAAAAAAGAAAAT------ 504
             |||      ||||||          :::       |||   :::|||
Db   3839 ThrLysGlySerHisValCysSerCysAlaLysAsnPheMetLysThrAspAsnMetCys 3858

Qy    505 ----------------------------------GCAGACGAGAACCAACTCTGCATC 528
                                            ||||||:::|||::::::
Db   3859 LysAlaGluGlySerGluHisGlnIleLeuTyrIleAlaAspAspAsnLysIleArgSer 3878
```

```
Qy      529 GTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGAC 588
            :::  ||||||       |||:::  |||
Db     3879 MetTyrProPheAsnProAsnSerAlaTyr------------------------------ 3888

Qy      589 GTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGAC--- 645
                                     |||:::            |||  ||||||  :::
Db     3889 -----------------GluProAla------------PheGlnGlyAspGluAsnVal 3898

Qy      646 ---------------------------------------------------AACTGGATG 654
                                                                ||||||
Db     3899 ArgIleAspAlaMetAspIleTyrValLysGlyAsnLysIleTyrTrpThrAsnTrp--- 3917

Qy      655 TGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGT------------ 702
                     |||||||||      |||          :::   |||
Db     3918 ---------HisThrGlyArgIleSerTyrCysGluLeuProAlaSerSerAlaAlaSer 3934

Qy      703 ---GCCGACAACCGG--------------------------------------------- 714
               |||  ||||||
Db     3935 ThrAlaSerAsnArgAsnArgArgGlnIleAspGlyGlyValThrHisLeuAsnIleSer 3954

Qy      714 ------------------------------------------------------------ 714

Db     3955 GlyLeuLysMetProArgGlyIleAlaValAspTrpValAlaGlyAsnIleTyrTrpThr 3974

Qy      715 -----------------------CCGGCCGAATACAGCAAGGACAATAAACCCTATAAG 750
                                    |||:::          :::|||:::
Db     3975 AspSerGlyArgAspValIleGluValAlaGlnMetLysGlyGluAsnArg--------- 3991

Qy      751 CCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACC 810
                                       :::   |||        :::    :::|||:::
Db     3992 --------------LysThrLeuIleSerGlyMetIleAspGluProHisAlaIleVal 4006

Qy      811 ATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATC 870
            :::      |||                  |||   :::   ||||||
Db     4007 ValAsp---ProLeuArgGlyThrMetTyrTrpSerAspTrpGly--------------- 4020

Qy      871 GAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAA 930
                     |||:::|||:::||||||          :::      |||      :::|||
Db     4021 ---------AsnHisProLysIleGluThrAlaAlaMet---AspGlyThrLeuArgGlu 4036

Qy      931 GCCATGAGCGCAGATCAGGCTACCCGTATCAAATAT------------GCCAGCAAGTAT 978
              :::      |||            |||:::::::              |||     |||
Db     4037 ThrLeuValGlnAsp---------AsnIleGlnTrpProThrGlyLeuAlaValAspTyr 4053

Qy      979 GCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTT 1038
                              ||||||  :::                      |||:::|||
Db     4054 HisAsnGluArgLeuTyrTrpAlaAsp-----------------------AlaLysLeu 4065

Qy     1039 GACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAAC 1098
               |||||||||                                 ||||||   |||
Db     4066 SerValIleGlySer--------------------------------IleArgLeuAsn 4074

Qy     1099 GGC--------------------------------------------------------- 1101
            |||
Db     4075 GlyThrAspProValValAlaIleAspAsnLysLysGlyLeuSerHisProPheSerIle 4094

Qy     1102 ---------AAGAGTGCTGTCTATGGC------GATGTATTGTCTTCTCTCGAAAAGGCT 1146
                     :::      :::||||||         :::    :::   :::   |||
Db     4095 AspIlePheGluAspTyrIleTyrGlyValThrTyrIleAsnAsnArgIlePheLysIle 4114

Qy     1147 TATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGT 1206
            :::|||   |||   |||:::            :::|||   |||:::        |||
Db     4115 HisLysPheGlyHisLysSer---------ValThrAsnLeuThrSerGlyLeuAsnHis 4131

Qy     1207 GGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCT 1266
              |||:::||||||      :::   |||:::               |||||||||
Db     4132 AlaThrAspValValLeuTyrHisGlnTyrLysGlnProGluValThrAsnPro------ 4149

Qy     1267 CATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTC------------ 1314
                                     |||   |||      :::::::|||
Db     4150 --------------------CysAspArgLysLysCysGluTrpLeuCysLeuLeuSer 4162

Qy     1315 CCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGC-------- 1366
             ||||||      |||      ||||||            |||      |||
Db     4163 ProSer-GlyProValCysThrCysProAsnGlyLysArgLeuAspAsnGlyThrCysVa 4182

Qy     1367 ----GTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTC 1422
              :::      ||||||:::|||    ::: :::            :::      ::::::
Db     4182 lLeuIleProSerProThrAlaSerAlaVal-ValProThrThrAspThrCysAspLeuV 4202

Qy     1423 AAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGC 1482
                     :::            |||   |||      |||                    :
Db     4202 alCysLeuAsnGlyGlySerCysPheLeuAsnAlaArg--------------------L 4215

Qy     1483 GACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAA---------- 1532
              ::|||      |||      |||            |||:::   ||||||
Db     4215 ysGlnAlaLysCysArgCysGlnProArgTyrAsnGlyGluArgCysGlnIleAsnGlnC 4235

Qy     1533 -----GGCTATCGAGAAAGATCCGGC-----------AGTAGAGCTTTCCAAGAGCGTA 1575
                     |||   :::      |||            |||   :::
Db     4235 ysSerAspTyrCysGlnAsnGlyGlyLeuCysThrAlaSerProSerGlyMetProThrC 4255

Qy     1576 ATAGCTGCTGCTCGCGCTATTCAGGC---CGATGCGATGGCCAA---TGC------CTAT 1623
               |||   :::   ::::::|||   |||||||||   |||   |||        |
Db     4255 ysArgCysProThrGlyPheThrGlySerArgCysAspGlnGlnValCysThrAsnTyrC 4275

Qy     1624 GCCATTGAGAAGGGCAAGCGTCTTTT------------------CTTTGCCGGTTTGCGT 1665
             |||||      |||      :::                          ||||||
Db     4275 ysHisAsnAsnGlySerCysThrValAsnGlnGlyAsnGlnProAsnCysArg------- 4292

Qy     1666 GAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTAC 1725
                     |||   |||         :::|||||||||           :::
Db     4293 -----CysProProThrPheIleGlyAspArgCysGlnTyrGlnGlnCysPheAsnTyrC 4311

Qy     1726 GGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGAC 1781
                  :::   |||:::                       |||:::|||   |||
Db     4311 ysGluAsnAsnGlyVal--------------------CysGlnMetSerArgAsp 4322
```

```
RESULT 6
YM96_YEAST
ID   YM96_YEAST     STANDARD;      PRT;  1140 AA.
AC   Q04893;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 113.1 kDa protein in PRE5-FET4 intergenic region.
GN   YMR317W OR YM9924.09.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z54141; CAA90835.1; -.
DR   SGD; S0004936; YMR317W.
KW   Hypothetical protein; Repeat.
SQ   SEQUENCE   1140 AA;  113070 MW;  0153EBCA24FE5427 CRC64;

Alignment Scores:
Pred. No.:              0.195          Length:          1140
Score:                  137.50         Matches:         131
Percent Similarity:     36.73%         Conservative:    96
```

```
Best Local Similarity:  21.20%      Mismatches:   246
Query Match:            3.60%       Indels:       145
DB:                     1           Gaps:         21

US-10-008-355-1 (1-2139) x YM96_YEAST (1-1140)

Qy     533 AGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTC 592
           |||   |||   |||   :::::|||   |||   :::     :::|||   |||
Db     233 SerThrSerSerProAlaSerSerThrIleSerGluThrLeuProPheSerSerThrIle 252

Qy     593 GTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGA 652
           :::         |||            |||||||||||||||         |||
Db     253 LeuSerIleThrSerSerPro---------ValSerSerGluAlaProSerAlaThrSer 269

Qy     653 TGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATG-----------CCG 700
                 :::   |||:::|||:::   |||:::   :::             |||
Db     270 SerSerValSerSerGluAlaSerSerSerThrSerSerSerValSerSerGluAlaPro 289

Qy     701 GTGCCGACAACCGGC---------------CGGCCGAATACAGCAAGGACAATA------ 739
           :::   |||:::                    |||::::::      ::::::
Db     290 LeuAlaThrSerSerValValSerSerGluAlaProSerSerThrSerSerValValSer 309

Qy     740 ------AACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTG 793
                 |||   |||      :::|||:::                   ::::::   :::
Db     310 SerGluAlaProSerSerThrSerSerSerVal---------------SerSerGluIle 324

Qy     794 ACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTT--- 850
           :::::            |||||||||            :::|||||||||   :::
Db     325 SerSerThrThrSerSerSerValSerSerGluAlaProLeuAlaThrSerSerValVal 344

Qy     851 -----------------------------GGGGTGTGGAAGATCGTATCGAAAACGAGA 880
                                        |||         :::|||   |||
Db     345 SerSerGluAlaProSerSerThrSerSerSerValSerSerGluIleSerSerThrThr 364

Qy     881 ACAATCCTCGTATCGAAG---------TTCGCGGTATCAAGCAAGGCATCTGGAAGGAAG 931
           :::   ||||||                |||   ||||||      |||
Db     365 SerSerSerValSerSerGluAlaProLeuAlaThrSerSerValValSerSerGluAla 384

Qy     932 CCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTA 991
           |||   :::                     |||      |||:::|||         :::
Db     385 ProSerSerThrSerSerSerValSerSerGluAlaProSerSerThrSerSerSerVal 404

Qy     992 ACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTC 1051
           :::          |||      ::::::||||||      ::::::
Db     405 SerSerGluAlaProSerSerThrSerSerSerValSerSerGluIleSerSerThrLys 424

Qy    1052 GTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTG 1111
              ||||||         |||   |||         |||         :::
Db     425 SerSerValMetSerSerGluValSerSerAlaThrSerSerLeuValSerSerGluAla 444

Qy    1112 TCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACC 1171
                 |||:::      |||   |||   ||||||         :::
Db     445 ProSerAlaIleSerSerLeuAlaSerSerArgLeuPheSerSerLysAsnThrSerVal 464

Qy    1172 GTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCAC 1231
                 |||:::            |||||||||
Db     465 ThrSerThrLeuValAlaThrGluAlaSerSerValThrSer------------------ 478

Qy    1232 AGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCG-----------GTATCC 1279
                                         |||   |||          ::::::
Db     479 -----------------------------SerLeuArgProSerSerGluThrLeuAla 488

Qy    1280 TCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGC 1339
           ||||||   :::      |||:::   :::|||         |||||||||
Db     489 SerAsnSerIleIleGluSerSerLeuSerThrGlyTyrAsnSerThrVal--------- 505

Qy    1340 CCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCA 1399
                               :::      :::      ||||||   :::   |||
Db     506 --------------------SerThrThrThrSerAlaAlaSerSerThrLeuGlySer 518

Qy    1400 AGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCG------ 1453
           ::::::   |||:::   :::      ||||||   ::::::      ||||||
Db     519 LysValSerSerSerAsnSerArgMetAlaThrSerLysThrSerSerThrSerSerAsp 538

Qy    1454 ------------------------------------TATTCGACAAGAGTGTGGTTC 1474
                                                   :::|||
Db     539 LeuSerLysSerSerValIlePheGlyAsnSerSerThrValThrThrSerProSerAla 558

Qy    1475 CTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGG---------ACAAGGAAAAGT 1525
              |||:::   ::::::   |||      |||   |||            |||
Db     559 SerIleSerLeuThrAlaSerProLeuProSerValTrpSerAspIleThrSerSerGlu 578

Qy    1526 TTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTG 1585
                    :::|||      :::         |||   |||
Db     579 AlaSerSerIleSerSerAsnLeuAlaSerSerSerAlaProSerAsp----------- 594

Qy    1586 CTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTC 1645
                                                             :::::::::
Db     595 ------------------------------------------AsnAsnSerThrIle 599

Qy    1646 TTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACT 1705
              |||   :::   |||      |||   :::   :::      ::::::   :::
Db     600 AlaSerAlaSerLeuIleValThrLysThrLysAsnSerValValSerSerIleValSer 619

Qy    1706 TCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGT 1765
           |||   :::   ::::::   :::   :::   |||      |||
Db     620 Ser------IleThrSerSerGluThrThrAsnGluSerAsnLeuAlaThr---SerSer 636

Qy    1766 ACAACTATCATACGACAG---------GCAAGGGCGTATTGGAGAAGCAGGATCCTAAGA 1816
           |||:::::::::               |||||||:::         |||
Db     637 ThrSerLeuLeuSerAsnLysAlaThrAlaArgSerLeuSerThrSerAsnAlaThrSer 656

Qy    1817 GCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCT-------------------- 1855
           |||   :::::|||               ||||||   |||
Db     657 AlaSerAsnValProThrGlyThrPheSerSerMetSerSerHisThrSerValIleThr 676

Qy    1856 ---------TCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCG 1906
                    |||               :::|||:::      ||||||   |||   |||
Db     677 ProGlyPheSerThrSerSerAlaSerLeuAlaIleAsnSerThrValValSerSerSer 696

Qy    1907 CTT---------TCCTATCGAACA--------------ACGACATCACGGGCGGTAACT 1942
           |||         |||:::   |||              |||||||||      ||||||
Db     697 LeuAlaGlyTyrSerPheSerThrProGluSerSerProThrThrSerThrLeuValThr 716

Qy    1943 CCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTC----------------- 1984
                 ||||||   ::::::   :::         |||
Db     717 SerGluAlaProSerThrValSerSerMetThrThrSerAlaProPheIleAsnAsnSer 736

Qy    1985 -------------------------------TTGCTTTCGATGGCAACTGGGAAGCTA 2011
                                              ::::::   :::|||      :::
Db     737 ThrSerAlaArgProSerProSerThrAlaSerPheIleThrGluSerThrSerSerIle 756

Qy    2012 TGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCT 2071
              |||         :::|||      :::   |||:::         :::|||
Db     757 SerSerValProLeuAlaSerGlyAspValThrSerSerLeuAlaAlaHisAsnLeuThr 776

Qy    2072 ACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
           |||||||||:::         |||:::      ||||||   :::   |||||||||
Db     777 ThrPheSerAlaProSerThrSer------SerAlaGlnLeuValSerLysSer 792

RESULT 7
PT1_CHLTR
ID   PT1_CHLTR      STANDARD;      PRT;   571 AA.
AC   O84340;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9)
DE   (Phosphotransferase system, enzyme I).
GN   PTSI OR CT336.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/Cx;
```

```
RX   MEDLINE=99000809; PubMed=9784136;
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. ENZYME I TRANSFERS THE PHOSPHORYL GROUP FROM
CC       PHOSPHOENOLPYRUVATE (PEP) TO THE PHOSPHORYL CARRIER PROTEIN (HPR).
CC       ENZYME I IS COMMON TO ALL PTS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + protein L-histidine =
CC       pyruvate + protein N(pi)-phospho-L-histidine.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE001306; AAC67931.1; -.
DR   InterPro; IPR000121; PEP_utilizers.
DR   Pfam; PF00391; PEP-utilizers; 1.
DR   Pfam; PF02896; PEP-utilizers_C; 1.
DR   ProDom; PD000940; PEP_utilizers; 1.
DR   PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; FALSE_NEG.
DR   PROSITE; PS00742; PEP_ENZYMES_2; FALSE_NEG.
KW   Phosphotransferase system; Transferase; Kinase; Sugar transport;
KW   Phosphorylation; Complete proteome.
FT   MOD_RES     207     207       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   571 AA;  63786 MW;  73EBFEEE7C457067 CRC64;

Alignment Scores:
Pred. No.:              0.362        Length:        571
Score:                  132.00       Matches:       115
Percent Similarity:     35.37%       Conservative:  88
Best Local Similarity:  20.03%       Mismatches:    186
Query Match:            3.33%        Indels:        185
DB:                     1            Gaps:          24

US-10-008-355-1 (1-2139) x PT1_CHLTR (1-571)

Qy   1635 CTTCTCAATGGCATAGGCATTGGCCATCGCATCGGCCTGAATAGCGCGAGCAGCAGCTAT 1576
          ::::::   |||||||||:::|||         :::
Db     22 IleValProGlyIleGlyLeuGlyLysAlaLeu-------------------------- 32

Qy   1575 TACGCTCTTGGAAAGCTCTACTGCCGGATCTTTCTCGATAGCCTTGGCAAACTTTTCCTT 1516
                |||||||||             ||||||         :::|||    :::      :::
Db     33 ---LeuLeuGlyLysSerSerLeuArgIleArgGluLeuThrLeuProGlnGluGluVal 51

Qy   1515 GTCCATGGACTTGAGCAT-----------------------GGCATGGAACTTGTCGCT 1480
             |||   :::                                     :::|||    |||
Db     52 GluHisGluIleSerArgTyrTyrLysAlaLeuLysArgSerArgSerAspLeuAlaAla 71

Qy   1479 ATAAGGAACCACACTCTTGTCGAA---------TACGAAGTCTGCATACTTCTTCGTGTC 1429
          :::                :::              |||:::      :::|||:::
Db     72 LeuGluLysGluAlaLysGlyLysGlnGlyTyrGlnGluIleAlaSerIleLeuGlnAla 91

Qy   1428 GCCTTTGAATTTCTTGTCGATTACATTCTT---GAATATATCGGGGAGCTTGTCGGCAGG 1372
                |||   :::   |||   :::|||   |||   :::   :::   :::   :::
Db     92 HisLeuGluIleIleLysAspProLeuLeuThrGluGluValValLysThrIleArgLys 111

Qy   1371 GATACGCCGGCGTACAATATCGAGCATGGCGGGCAGCACCTTACGGTCGAGCGAGGGGAG 1312
          |||                                            :::      |||
Db    112 AspArg------------------------------------------LysAsnAlaGlu 117

Qy   1311 GTAGTCTTTGTACTTGTCGTCAAGCGATTTGAGGATACCGGCATGAGCATCAGG------1258
                ||||||      |||:::      |||   :::      ::::::::

Db    118 PheValPheSerSerValMetGlyGluIleGluLysSerLeuCysAlaValGlnLysThr 137

Qy   1257 -------------------------------------ATTTGTAGCCAATGCGTTGGC 1237
                                                  |||   ::::::   :::|||
Db    138 ThrAlaThrArgValAspArgValGlnAspIleHisAspIleSerAsnArgValIleGly 157

Qy   1236 AAACTGTGC------AAAACGAACCACCTCGGTACCACCGAAGAGCGTCTC--------- 1192
             ||||||           :::   ||||||       ::::::    |||
Db    158 HisLeuCysCysGlnHisLysSerSerLeuGlyGluPheAspGlnAsnLeuIleValPhe 177

Qy   1192 ------------------------------------------------------------ 1192

Db    178 SerGluGluLeuThrProSerGluAlaAlaAsnAlaAsnProGluTyrIleArgGlyPhe 197

Qy   1191 --------------GCTCAAATAAGTCATCTCACGGTTGGCCTTGGCTCCTTCCTTATA 1147
                         |||:::   ||||||       :::                 :::
Db    198 ValSerLeuGluGlyAlaLysThrSerHisThrAlaIle------------------Val 211

Qy   1146 AGCCTTTTCGAGAGAAGACAATACATCGCC------------------------------ 1117
          ||||||    :::          |||:::|||
Db    212 SerLeuAlaLysAsnIleProTyrValAlaAsnPheThrThrGluLeuTrpAspThrIle 231

Qy   1116 ------ATAGACAGCACTCTTGCCGTTCTTACGGATCCAGTCTGCGAATGCTCTTTC--- 1066
                         ||||||    ::::::               |||    :::|||
Db    232 LysGluPheSerGlyThrLeuValLeuIleAsnGlyAspLysGlyGluIleThrPheAsn 251

Qy   1065 ------CTCGGCACGCTTACGACCTATCACGTCAAG-----------------ACGAGC 1030
                |||   |||:::   |||||||:::   |||                  |||
Db    252 ProGlnLeuSerThrIleGlnThrTyrTyrArgLysGlnAlaSerValSerValThrVal 271

Qy   1029 GAGACCGCGGTTCATACCGATCGAATTCTTCCAATAGTTAGCACTCTGAGCATACTTGCT 970
                  |||   |||   :::   ||||||:::::|||  |||    |||
Db    272 ProValGlnValGlnThrGlyLysAsnLeuProLeuIleSer-LeuSerAla-------- 288

Qy    969 GGCATATTTGATACGGGTAGCCTGATCTGCGCTCATGGCTTCCTTCCAGATGCCTTGCTT 910
                                      ::::::::::::   ::::::|||    ::
Db    289 --------------------------GlnIleValSerThrGluGluLeuProMetIl 299

Qy    909 G---------ATACCGCGAACTTCGATACGAGGATTGTTCTCGTTTTCGATACGATCTTC 859
          :              |||   |||||||:::   |||||||||
Db    299 eGluArgGluSerProGlyThrSerVal---GlyLeuPhe------------------- 311

Qy    858 CACACCCCAAGAAGTGAGGTAGCGATCCGTACTGCCCGGGAAACCGATGGTCATGGCATA 799
                                                       :::    ||||||
Db    312 ---------------------------------------ArgSerGluPheMetAlaPh 318

Qy    798 GTCGTCAGCCTTGTAGCCTTGCATGGATACGGCAGCGAAGTAAACGGGCTTATAGGGTTT 739
           |||          ||||||::::::   |||      :::   |||       |||
Db    318 eSerLeuGlyArgLeuProCysValGlu-GluGlnAlaAspGlnTyrAlaGlnLeuValG 338

Qy    738 AT---------------------------------------------------------T 736
                                                                     |
Db    338 lnPheGlnCysSerAspIleHisValLeuArgLeuPheAspPheGlyGluAspLysGluC 358

Qy    735 GTCCTTGCTGTATTCGGCCGGCCGGTTGTCGGCACCGGCATACACGCGGAATACGCTGAA 676
          ||||||||                               ||||||    |||    :::    |||
Db    358 ysProCys-----------IleSerSerSerHisArgSerValArgTrpLeu-LeuGlu 373

Qy    675 GTCGCCCGTGTGACGCGGCCACAT---CCAGTTGTCCGTATCGCCTCCGAACTTACCTAC 619
             :::|||   :::                 :::    |||   |||
Db    374 GlnGluLysValLeuLysGluGlnLeuGlnAlaIleAlaIleValSer------------ 389

Qy    618 AGAGCTGGGAGGAGCAAATACCATACGAACGTCCTTGAATACATCGTAGACGATGAGGAA 559
          |||   ||||||   |||                  ::::::    ::::::|||    |||
Db    390 ArgIleGlyArgLeuLysVal---------LeuIleProGlyValIleAspAlaSerGlu 406

Qy    558 GTATTCGTTGTTGGAATAGAAAGGCTCTACGATGCAGAGTTGGTTCTCGTCTGCATTTTC 499
          :::   :::|||       :::|||||||:::          |||:::    |||:::
Db    407 IleAlaLeuVal------LysArgLeuPheGlnGluGluIleArgLeuLeuLysGlyIle 424

Qy    498 TTTTTTGGCCAGTTCTTGGCATACCTCCTGAGCTTTGCGCAGACGCTCCATCTCGTCAGT 439
                                 ::::::             :::    ::::::
Db    425 SerGluAsnIleLeuTrpGlySerMetIleGluIleProSerAlaValTrpMetIleGlu 444
```

```
Qy     438 GATACCCTTGAGCTGTCCTTCTACCTTGTCCGTTACCTTCACGATCTTGCGCAGATACTT 379
           :::   |||:::   |||   ::::::             :::||||||||||||
Db     445 GluIleLeuGlnGluSerSerPheValAlaLeuGlyThrAsnAspLeuAlaGlnTyrThr 464

Qy     378 CACGGAAAGACCCGGAATCGGAAGCTCCTCACCCATCGTGCGAGAAACGAAACCATCGCG 319
              |||   :::|||   |||   ||||||       |||:::|||            :::
Db     465 LeuGlyThrSerArgGluArgSerLeuLeuGlyGluArgSerArgValProHisProSer 484

Qy     318 CAGATAGTCGTGATCCACCGTGCTTTG------GCTCTGGATAGCACCGTATCCGCAGTG 265
              |||   :::||||||      :::      |||          |||
Db     485 ValIleArgMetIleHisHisValValGluGlnAlaLysGlnLysAsnValProValSer 504

Qy     264 GTGGTTGGTAAA---------GATCAGGCCCTGATC 238
            |||   |||:::          |||   ||||||:::
Db     505 ValCysGlyGluMetAlaGlyAspProAlaLeuLeu 516

RESULT 8
AGA1_YEAST
ID   AGA1_YEAST     STANDARD;      PRT;   725 AA.
AC   P32323;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   A-agglutinin attachment subunit precursor.
GN   AGA1 OR YNR044W OR N3431.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91304412; PubMed=2072914;
RA   Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;
RT   "The AGA1 product is involved in cell surface attachment of the
RT   Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin.";
RL   Mol. Cell. Biol. 11:4196-4206(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Pohl T.M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
CC       SUBUNIT. S.CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
CC       CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
CC       RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
CC       AGGREGATION DURING MATING.
CC   -!- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
CC       A CORE SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M60590; AAA34382.1; -.
DR   EMBL; Z71659; CAA96325.1; -.
DR   PIR; S17031; S17031.
DR   PIR; A41258; A41258.
DR   SGD; S0005327; AGA1.
KW   Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;
KW   Pheromone response.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    725       A-AGGLUTININ ATTACHMENT SUBUNIT.
FT   DOMAIN       53    493       2 X APPROXIMATE REPEATS.
FT   REPEAT       53    149       1-1.
FT   REPEAT      395    493       1-2.
FT   DOMAIN      182    307       18 X APPROXIMATE TANDEM REPEATS, SER/THR-
FT                                RICH.
FT   REPEAT      182    188       2-1.
FT   REPEAT      189    195       2-2.
FT   REPEAT      196    202       2-3.
FT   REPEAT      203    209       2-4.
FT   REPEAT      210    216       2-5.
FT   REPEAT      217    223       2-6.
FT   REPEAT      224    230       2-7.
FT   REPEAT      231    237       2-8.
FT   REPEAT      238    244       2-9.
FT   REPEAT      245    251       2-10.
FT   REPEAT      252    258       2-11.
FT   REPEAT      259    265       2-12.
FT   REPEAT      266    272       2-13.
FT   REPEAT      273    279       2-14.
FT   REPEAT      280    286       2-15.
FT   REPEAT      287    293       2-16.
FT   REPEAT      294    300       2-17.
FT   REPEAT      301    307       2-18.
SQ   SEQUENCE   725 AA;  73353 MW;  70420C853B0B01F8 CRC64;

Alignment Scores:
Pred. No.:              0.779           Length:         725
Score:                  127.50          Matches:        130
Percent Similarity:     38.73%          Conservative:   83
Best Local Similarity:  23.64%          Mismatches:     200
Query Match:            3.34%           Indels:         138
DB:                     1               Gaps:           22

US-10-008-355-1 (1-2139) x AGA1_YEAST (1-725)

Qy     509 ACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCG 568
           |||   |||:::|||:::|||   |||   |||         :::|||:::||||||||||||
Db     238 ThrSerThrSerSerSerSerThrSerThrSerGlnSerSerThrSerThrSerSerSer 257

Qy     569 TCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGT 628
           ||||||      ||| - :::                          :::      |||
Db     258 SerThrSerThrSerProSer----------------------SerThrSerThrSer 269

Qy     629 TCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCG-----------ACT 676
           |||   :::      |||:::          :::   |||   |||          |||
Db     270 SerSerSerThrSerThrSerProSerSerLysSerThrSerAlaSerSerThrSerThr 289

Qy     677 TCAGCGTATTCC---GCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGG 733
           |||:::||||||   :::      |||      |||:::      |||      |||
Db     290 SerSerTyrSerThrSerThrSerProSerLeuThrSerSerSerProThrLeuAlaSer 309

Qy     734 ACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTG 793
           |||            ||||||   ::::::
Db     310 Thr------------SerProSerSerThr----------------------------- 315

Qy     794 ACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGG 853
                                |||:::|||            :::||||||   ||||||
Db     316 -----------------SerIleSerSerThrPheThrAspSerThrSerSerLeuGly 329

Qy     854 GTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGC 913
                 |||   |||   :::   :::         ||||||   ::::::   |||:::
Db     330 Ser---SerIleAlaSerSerSerThrSer------ValSerLeuTyrSerProSerThr 346

Qy     914 AAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCA 973
                                                  ||||||   ::::::|||:::
Db     347 -------------------------------------ProValTyrSerValProSer 353

Qy     974 AGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCG---GTATGAACCGCGGTCTCG 1030
           :::         |||   |||                  |||   |||   |||   ||||||
Db     354 ThrSerSerAsnValAlaThrProSerMetThrSerSerThrValGluThrThrValSer 373

Qy    1031 CTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCC 1090
                                  |||         |||:::               |||
Db     374 ----------------------SerGlnSerSerSerGluTyrIleThrLysSerSer 385

Qy    1091 GTAAGAACGGC------------AAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTC 1137
           :::   |||                 :::   |||         :::|||   :::
```

```
Db     386 IleSerThrThrIleProSerPheSerMetSerThrTyrPheThrThrValSerGlyVal 405

Qy    1138 GAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAG---------------ATGACT 1182
                    |||                 :::   |||              :::|||
Db     406 ThrThrMetTyrThrThrTrpCysProTyrSerSerGluSerGluThrSerThrLeuThr 425

Qy    1183 TATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAAC 1242
              :::   ||||||:::                  |||
Db     426 SerMetHisGluThrValThrThrAspAlaThrValCysThrHisGluSerCysMetPro 445

Qy    1243 GCATTGGCTACAAATCCTGATGCTC------ATGCCGGTATCCTCAAATCGCTTGACGAC 1296
           :::      |||:::  |||:::        :::   :::|||:::       :::   ||
Db     446 SerGlnThrThrSer-LeuIleThrSerSerIleLysMetSerThrLysAsnValAlaTh 465

Qy    1297 AAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGC------TGCCCGCCATGCTC 1350
              ||||    |||::::::        |||:::      |||      |||      ||
Db     465 rSerValSerThrSerThrValGluSerSerTyrAlaCysSerThrCysAlaGluThrSe 485

Qy    1351 GATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATC 1410
           |       |||:::    |||            :::||||||   :::       :::    ||
Db     485 rHisSerTyrSerSerValGlnThrAlaSerSerSerSerValThrGlnGlnThrThrSe 505

Qy    1411 GACAAGA------------------------------------AATTCAAAGGCGAC 1431
           ||||:::                                    |||    |||||
Db     505 rThrLysSerTrpValSerSerMetThrThrSerAspGluAspPheAsnLysHisAlaTh 525

Qy    1432 ACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTC 1491
           |    :::             ||||||      |||   |||    :::       |||
Db     525 rGlyLysTyrHisValThrSerSerGlyThr--SerThrIleSerThrSerValSerGlu 544

Qy    1492 CATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCG 1551
                          |||:::|||    |||                  ::::::
Db     545 AlaThrSerThrSerSerIleAspSerGluSer---------------GlnGluGlnSer 559

Qy    1552 GCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATG 1611
           :::        ||||||   ||||||::::::::::::    ::::::     |||
Db     560 SerHisLeuLeuSerThrSerValLeuSerSerSer---SerLeuSerAlaThrLeuSer 578

Qy    1612 GCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTT----------- 1660
           :::::::::        |||        :::||||||   ||||||    |||
Db.    579 SerAspSerThrIle-LeuLeuPheSerSerValSerSerLeuSerValGluGlnSerPr 598

Qy    1661 ------------TGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTC 1707
                       :::       |||     ::::::|||           ||||||||
Db     598 oValThrThrLeuGlnIleSerSerThrSerGluIleLeu---------GlnProThrSe 615

Qy    1708 ACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTAC 1767
           |                                                         ::
Db     615 r-----------------------------------------------------------Se 616

Qy    1768 AACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTT 1827
           :|||    |||              |||            |||          :::::::|||
Db     616 rThrAlaIleAlaThrIleSerAlaSerThrSerSerLeuSerAlaThrSerIleSerTh 636

Qy    1828 GCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAG 1887
             |||          :::   ||||||       :::              ::::::
Db     636 rProSerThrSerValGluSerThrIleGluSerSer------------SerLeuThrPr 652

Qy    1888 AACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGT 1947
             ||||||||||||||||   ||||||                           :::
Db     652 oThrValSerSerIlePheLeuSer---------------------------SerSerSe 663

Qy    1948 AGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAA 2007
             ||||||    |||::::::|||:::|||      :::      :::|||      :::
Db     663 rAlaProSerSerLeuGlnThrSerValThrThrThrGluValSerThrThrSerIleSe 683

Qy    2008 GCTATGAGTGGTGACATCGAGT 2029
             :::         |||||||||
Db     683 rIleGlnTyrGlnThrSerSer 690
```

RESULT 9
YS89_CAEEL

```
ID   YS89_CAEEL     STANDARD;      PRT;   3178 AA.
AC   Q09624; Q09625; Q969D4;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein ZK945.9 in chromosome II.
GN   ZK945.9/ZK945.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Wilkinson-Sproat J.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
RA   Durbin R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z48544; CAB70192.1; -.
DR   EMBL; Z48582; CAB70192.1; JOINED.
DR   EMBL; Z48582; CAB70201.1; -.
DR   EMBL; Z48544; CAB70201.1; JOINED.
DR   WormPep; ZK945.9; CE25697.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR001024; Lipoxygenase_LH2.
DR   InterPro; IPR000636; M+channel_nlg.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00520; ion_trans; 1.
DR   Pfam; PF01477; PLAT; 1.
DR   Pfam; PF01825; GPS; 1.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00308; LH2; 1.
KW   Hypothetical protein; Transmembrane.
FT   DOMAIN      266   1196       SER/THR-RICH.
FT   DOMAIN     1105   1241       GLY/SER-RICH.
FT   DOMAIN     2071   2120       GPS.
FT   TRANSMEM     13     30       POTENTIAL.
FT   TRANSMEM     51     73       POTENTIAL.
FT   TRANSMEM   2139   2161       POTENTIAL.
FT   TRANSMEM   2348   2367       POTENTIAL.
FT   TRANSMEM   2390   2412       POTENTIAL.
FT   TRANSMEM   2451   2468       POTENTIAL.
FT   TRANSMEM   2483   2505       POTENTIAL.
FT   TRANSMEM   2567   2589       POTENTIAL.
FT   TRANSMEM   2836   2858       POTENTIAL.
FT   TRANSMEM   2939   2961       POTENTIAL.
FT   TRANSMEM   2976   2998       POTENTIAL.
FT   TRANSMEM   3038   3060       POTENTIAL.
SQ   SEQUENCE   3178 AA;  344726 MW;  F8239436D03666CD CRC64;
```

```
Alignment Scores:
Pred. No.:              1.47          Length:         3178
Score:                  126.50        Matches:        148
Percent Similarity:     31.05%        Conservative:   70
Best Local Similarity:  21.08%        Mismatches:     227
Query Match:            3.31%         Indels:         258
DB:                     1             Gaps:           29
```

US-10-008-355-1 (1-2139) x YS89_CAEEL (1-3178)

```
Qy     346 GGTGAGGAGCTTCCGATTCCGGGT--------------------------CTTTCCGTG 378
```

```
               |||   |||:::    :::    |||                            ::::::
Db      186 GlyAsnGluMetAlaLeuLeuGlyTyrArgGluLysCysGluSerGlyGluIleAsnGlu 205

Qy      379 AAGTATCTGCGCAAGATCGTGAAG----------------------------GTAACG 408
            :::|||   |||:::::   |||                            ::::::
Db      206 GluTyrAlaArgArgMetCysLysArgProTyrArgSerGluLysSerThrAlaIleSer 225

Qy      409 GACAAG-----------GTAGAAGGACAG---CTCAAGGGTATCACTGACGAGATGGAG 453
            |||                 :::||||||   |||||||||:::       :::
Db      226 AspSerGlnGlyValTyrTyrAspGlyGlnValLeuLysGlyValArgAlaLysGlnPhe 245

Qy      454 CGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAG 513
               :::|||   :::                        |||::::::     |||
Db      246 SerMetArgThrSerGlySerProThrLeuArgArgMetLysArgAspAlaGlyAsp-As 265

Qy      514 AACC--------------AACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATAC 558
             |||                |||   |||   |||   :::   |||||||||   ||
Db      265 nThrCysAspTyrThrIleGluSerThrSerThrSerThrThrThrProThrThrThrTh 285

Qy      559 TTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCT 618
            |   :::|||:::
Db      285 rValThrSerThr------------------------------------------------ 289

Qy      619 GTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTC 678
                |||:::|||                |||               ||||||   |||   ||
Db      290 ----ValThrSerThrThrThrValProThrSerThrSerThrValThrThrAlaMetSe 308

Qy      679 AGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAGGACAAT 738
             |                       :::|||   |||:::|||:::    |||||
Db      308 rThr-----------------------SerThrSerThrProSerThrSerThrThrIl 320

Qy      739 AAACCCTATAAGC---CCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGAC 795
            |           |||      |||||||||
Db      320 eGluSerThrSerThrThrPheThrSer------------------------------ 329

Qy      796 GACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGT 855
                                                  :::||||||
Db      330 ---------------------------------ThrAlaSerThrSer----------- 334

Qy      856 GTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAA 915
                       |||   |||   |||             ::::::|||
Db      335 ---------ThrSerSerThrSerThrThrGlnGlnSerSerSerThrIleThrSerSe 351

Qy      916 GGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCA----- 970
               |||                             :::|||   :::     |||
Db      351 rProSerSerThrThr------LeuSerThrSerIleProThrThrThrThrProGluIl 369

Qy      971 -GCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTA------TGAACCGC 1023
               |||   |||      |||                  |||   |||         |||
Db      369 eThrSerThrLeuSerSerLeuProAspAsnAlaIleCysSerTyrLeuAspGluThrTh 389

Qy     1024 GGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGAC 1083
               |||         |||      ::::::::               |||         ||
Db      389 rThrSerThrThrPheThrThrThrMetLeuThrSerThrThrThrGluGluProSerTh 409

Qy     1084 TGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAG 1143
            |   |||   |||   |||   |||   |||
Db      409 r---SerThrThrThrThrGluValThrSer------------------ThrSerSerTh 422

Qy     1144 GCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTC 1203
               :::          ||||||      |||:::
Db      422 rValThrThrThrGluProThrThrThrLeu----------------------------- 432

Qy     1204 GGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGAT 1263

Db      432 ------------------------------------------------------------ 432

Qy     1264 GCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCT---CG 1320
                                       |||   ||||||   ||||||   |||
Db      433 ------------------ThrThrSerThrAlaSerThrSerThrThrGluProSerTh 446

Qy     1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
             |||||||||                                         :::   |||::
Db      446 rSerThrVal----------------------------------ThrThrSerProSe 454

Qy     1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440
            ::::||||||:::   |||   :::   |||:::   :::|||       |||
Db      454 rThrSerProValThrSerThrValThrSerSerSerSerSerThrThr-------- 471

Qy     1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500
             :::   |||      ||||||                    :::||||||:::   |||
Db      472 -ValThrThrProThrSerThrGlu--------------SerThrSerThrSerProSe 486

Qy     1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAG----------------- 1542
             |||:::      |||      :::   |||      |||
Db      486 rSerThrValThrThrSerThrThrAlaProSerThrSerThrThrGlyProSerSerSe 506

Qy     1543 ----AAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCA 1598
                |||:::         |||   ||||||   :::   |||   :::   ||
Db      506 rSerSerThrProSerSerThrAlaSerSerSerValSerSerThrAlaSerSerThrGl 526

Qy     1599 GGCCGATGCGATGGCC-------------------------------------------- 1614
            |:::
Db      526 nSerSerThrSerThrGlnGlnSerSerThrThrThrLysSerGluThrThrThrSerSe 546

Qy     1615 ----------------AATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGG 1658
                           :::   |||   :::||||||          |||:::
Db      546 rAspGlyThrAsnProAspPheTyrPheValGluLysAlaThrThrThrPheTyrAspSe 566

Qy     1659 TTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTAT 1718
                                                 :::   |||   |||:::
Db      566 r-----------------------------------ThrSerValAsnLeuThrLeuAsnSe 575

Qy     1719 GAGCTACGGCTCCATCAAGGGATATGAA------------------------------ 1746
               |||   |||   ||||||:::
Db      575 rGlyLeuGlyIleIle---GlyTyrGlnThrSerIleGluCysThrSerProThrSerSe 594

Qy     1747 ---------------CCGCAGGACGGTGCCTGGTACAAC------------------ 1770
                            :::|||||||||   :::
Db      594 rAsnTyrValSerThrThrLysAspGlyAlaCysPheThrLysSerValSerMetProAr 614

Qy     1770 ------------------------------------------------------------ 1770

Db      614 gLeuGlyGlyThrTyrProAlaSerThrPheValGlyProGlyAsnTyrThrPheArgAl 634

Qy     1771 ---------------------------TATCATACGACAGGCAAGGGCGTAT------- 1795
                                       |||:::|||   |||      |||
Db      634 aThrMetThrThrAspAspLysLysValTyrTyrThr-TyrAlaAsnValTyrIleGlnG 654

Qy     1796 --TGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCT 1853
               :::   |||      :::   :::   |||              |||||||||
Db      654 luTyrSerSerThrThrIleGluSerGluSerSerThrSerAlaValAlaSerSerThrS 674

Qy     1854 CTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACG--------------GTCAGCT 1898
            |||||   |||   |||         :::            |||            ||||
Db      674 erSerThrProSerThrProSerSerThrLeuSerThrSerThrValThrGluProSerS 694

Qy     1899 CCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATT 1958
            ||          |||      |||||||||||         |||            |
Db      694 erThrArgSerSerAspSerThrThrThrSerAlaGlySerThrThrThrLeuGlnGluS 714

Qy     1959 CGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGG 2018
            ||      |||:::         |||          |||       |||
Db      714 erThrThrThrSerGluGluSerThrThrAspSerSerThrThrThrIleSerAspThrS 734

Qy     2019 TGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCT 2078
               |||:::||||||:::|||         |||:::   :::|||   :::::::|||   |
Db      734 erThrThrSerSerSerPro------SerSerThrThrAlaAspSerThrSerThrLeuS 752

Qy     2079 CT 2080
            ||
Db      752 er 752

RESULT 10
COLA_CLOPE
ID   COLA_CLOPE     STANDARD;      PRT;   1104 AA.
```

```
AC   P43153;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Microbial collagenase precursor (EC 3.4.24.3) (120 kDa collagenase).
GN   COLA OR CPE0173.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 87-113.
RC   STRAIN=NCIB 10662;
RX   MEDLINE=94110220; PubMed=8282691;
RA   Matsushita O., Yoshihara K., Katayama S.-I., Minami J., Okabe A.;
RT   "Purification and characterization of Clostridium perfringens 120-
RT   kilodalton collagenase and nucleotide sequence of the corresponding
RT   gene.";
RL   J. Bacteriol. 176:149-156(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN   [3]
RP   SEQUENCE OF 1073-1104 FROM N.A.
RC   STRAIN=NCIB 10662;
RA   Matsushita O.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Digestion of native collagen in the triple
CC       helical region at Xaa-|-Gly bonds. With synthetic peptides, a
CC       preference is shown for Gly at P3 and P1'; Pro and Ala at P2 and
CC       P2'; and hydroxyproline, Ala or Arg at P3'.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M9.
CC   -!- SIMILARITY: CONTAINS 1 PKD DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D13791; BAA02941.1; -.
DR   EMBL; AP003185; BAB79879.1; -.
DR   EMBL; D50309; BAA08848.1; -.
DR   MEROPS; M09.002; -.
DR   InterPro; IPR002169; Micollptase.
DR   InterPro; IPR000601; PKD_domain.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF00801; PKD; 1.
DR   Pfam; PF01752; Peptidase_M9; 1.
DR   PRINTS; PR00931; MICOLLPTASE.
DR   SMART; SM00089; PKD; 1.
DR   PROSITE; PS50093; PKD; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Zymogen; Signal; Complete proteome.
FT   SIGNAL        1     39       POTENTIAL.
FT   PROPEP       40     86
FT   CHAIN        87   1104       MICROBIAL COLLAGENASE.
FT   DOMAIN      774    862       PKD.
FT   METAL       502    502       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    503    503       BY SIMILARITY.
FT   METAL       506    506       ZINC (CATALYTIC) (BY SIMILARITY).
FT   CONFLICT     38     38       L -> F (IN REF. 1).
FT   CONFLICT    722    722       I -> M (IN REF. 1).
FT   CONFLICT    748    748       G -> E (IN REF. 1).
FT   CONFLICT    945    945       V -> E (IN REF. 1).
FT   CONFLICT    970    970       T -> A (IN REF. 1).
FT   CONFLICT    987    987       A -> E (IN REF. 1).
FT   CONFLICT   1098   1098       I -> T (IN REF. 1 AND 3).
SQ   SEQUENCE   1104 AA;  125935 MW;  F4B7377194ED021C CRC64;

Alignment Scores:
Pred. No.:              2.08            Length:         1104
Score:                  122.00          Matches:        141
Percent Similarity:     31.46%          Conservative:   105
Best Local Similarity:  18.03%          Mismatches:     278
Query Match:            3.19%           Indels:         258
DB:                     1               Gaps:           36

US-10-008-355-1 (1-2139) x COLA_CLOPE (1-1104)

Qy      85 TGGCTCCTCAACGAACTCAATCAGGAGAATCTG--------------GATCGAATGCGT 129
              |||         |||:::      :::   :::                |||:::
Db     419 TrpAlaSerLysGluValLysAlaGlnPheMetArgValValGlnAsnAspLysAlaLeu 438

Qy     130 GAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCC 189
              |||   |||            :::   ::::::      ::::::   |||
Db     439 GluGluGlyAsnProAspAspIleLeuThrValValIleTyrAsnSerProGluGluTyr 458

Qy     190 AATGCCGTGGTTATCTTCGGTGGCGGATGTACC---------GGTATCACAGTGTCCGAT 240
                          |||      |||      |||      ||||||   :::   :::
Db     459 LysLeuAsnArgIleIleAsnGlyPheSerThrAspAsnGlyGlyIleTyrIleGluAsn 478

Qy     241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGC---CAAAGC 297
                 |||      ||||||                     :::|||   :::   :::
Db     479 IleGlyThrPhePheThrTyrGluArgThrProGluGluSerIleTyrThrLeuGluGlu 498

Qy     298 ACGGTGGATCACGAC---------TATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGT 348
                       |||:::         ||||||:::      :::|||
Db     499 LeuPheArgHisGluPheThrHisTyrLeuGlnGlyArgTyrVal--------------- 513

Qy     349 GAGGAGCTTCCGATTCCGGGTCTT---------TCCGTGAAGTATCTGCGCAAGATCGTG 399
                       :::||||||:::                     |||         ::::::
Db     514 ------------ValProGlyMetTrpGlyGlnGlyGluPheTyrGlnGluGlyValLeu 529

Qy     400 AAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTG 459
                    ::::::                        :::         :::   :::
Db     530 ThrTrpTyrGluGluGlyThrAlaGluPhePheAlaGlySerThrArgThrAspGlyIle 549

Qy     460 CGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAA 519
              :::      :::   |||   |||   ||||||         |||         :::
Db     550 LysProArgLysSerValThrGlnGlyLeuAlaTyrAspArgAsnAsnArgMetSerLeu 569

Qy     520 CTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTA 579
                 ::::::            |||   :::   ::::::::
Db     570 TyrGlyValLeuHisAlaLysTyrGlySerTrpAspPheTyr------------------ 583

Qy     580 TTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGAT 639
                                                                :::|||
Db     584 ---------------------------------------------AsnTyrGlyPheAla 588

Qy     640 ACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCC 699
                 |||:::|||:::         :::   |||   |||
Db     589 LeuSerAsnTyrMetTyrAsnAsnAsnMetGlyMetPhe--------------------- 601

Qy     700 GGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTAC 759
                       |||:::         |||   |||:::|||
Db     602 ---------AsnLysMetThrAsnTyrIleLysAsnAsn--------------------- 611

Qy     760 TTCGCTGCCGTATCCATGCAAGGCTACAAG-----------------GCTGACGACTAT 801
                          :::   |||||||||                :::    ||||||
Db     612 -----------AspValSerGlyTyrLysAspTyrIleAlaSerMetSerSerAspTyr 627

Qy     802 GCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAA 861
              :::         :::               |||      |||
Db     628 GlyLeuAsnAspLysTyrGlnAspTyrMetAspSerLeuLeu------------------ 641

Qy     862 GATCGTATCGAAAACGAGAACAATCCTCGTATCGAA---------GTTCGCGGTATCAAG 912
           :::   |||:::|||   :::   |||   :::                |||   |||
```

```
Db    642 AsnAsnIleAspAsnLeuAspValProLeuValSerAspGluTyrValAsnGly------ 659

Qy    913 CAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGT---ATCAAATATGCC 969
                         ||||||         ::::::  |||       ||||||
Db    660 ------------HisGluAlaLysAspIleAsnGluIleThrAsnAspIleLysGluVal 675

Qy    970 AGCAAGTATGCTCAG---AGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGT 1026
          |||               |||:::|||      |||:::
Db    676 SerAsnIleLysAspLeuSerSerAsnValGluLysSerGlnPhePheThrThrTyrAsp 695

Qy   1027 CTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGG 1086
          :::            |||   ||||||   :::      ||||||            ||||||
Db    696 MetArgGlyThrTyrValGlyGlyArgSerGlnGlyGluGluAsn---------AspTrp 712

Qy   1087 ATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAG--- 1143
                   |||   |||               |||:::|||      |||  |||
Db    713 LysAspMetAsnSerLys---------LeuAsnAspIleLeuLysGluLeuSerLysLys 729

Qy   1144 --------------------------GCTTATAAGGAAGGAGCCAAGGCCAACCGTGAG 1176
                                     ||||||            |||   :::
Db    730 SerTrpAsnGlyTyrLysThrValThrAlaTyrPheValAsnHisLysValAspGlyAsn 749

Qy   1177 ATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTT 1236
                |||:::   :::   :::|||   |||
Db    750 GlyAsnTyrValTyrAspValValPheHisGlyMetAsnThr------------------ 763

Qy   1237 GCCAACGCATTGGCTACAAATCCTGATGCTCATGCC---------------GGTATCCTC 1281
                         ||||||   |||   |||                        ::::::
Db    764 ------------AspThrAsnThrAspValHisValAsnLysGluProLysAlaValIle 779

Qy   1282 AAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCC 1341
          ||||||                                    |||      |||
Db    780 LysSer------------------------------------AspSerSerVal------ 785

Qy   1342 GCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAG 1401
                      ||||||         |||                        |||
Db    786 ------------IleValGluGluGluIleAsn--------------------PheAsp 794

Qy   1402 AATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTAT------------------ 1443
                      |||         |||:::   |||   |||
Db    795 GlyThrGluSerLysAspGluAspGlyGluIleLysAlaTyrGluTrpAspPheGlyAsp 814

Qy   1444 ------------------------GCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTAT 1479
                                  |||         ::::::|||
Db    815 GlyGluLysSerAsnGluAlaLysAlaThrHisLysTyrAsnLys--------------- 829

Qy   1480 AGCGACAAGTTCCATGCCATGCTCAAGTCCATGGAC---------------------AAG 1518
          :::   ::::::         |||         |||                     :::
Db    830 ThrGlyGluTyrGluValLysLeuThrValThrAspAsnAsnGlyGlyIleAsnThrGlu 849

Qy   1519 GAAAAGTTTGCCAAGGCTATCGAGAAAGATCCG--------------------------- 1551
             |||      |||   :::|||      |||
Db    850 SerLysLysIleLysValValGluAspLysProValGluValIleAsnGluSerGluPro 869

Qy   1552 ------------------GCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCT 1593
                            |||   ::::::::||||||      :::
Db    870 AsnAsnAspPheGluLysAlaAsnGlnIleAlaLysSerAsnMetLeuValLysGlyThr 889

Qy   1594 ATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGT--------- 1644
          :::      :::      ::::::   |||      :::      |||:::
Db    890 LeuSerGluGluAspTyrSerAspLysTyrTyrPheAspValAlaLysLysGlyAsnVal 909

Qy   1645 ---------------------------------------------CTTTTCTTTGCCGGT 1659
                                                       |||:::      |||
Db    910 LysIleThrLeuAsnAsnLeuAsnSerValGlyIleThrTrpThrLeuTyrLysGluGly 929

Qy   1660 TTGCGTGAGATGTAC---------------------CCCGGACGTGCTCTGCCGAGCGAT 1698
                      |||                        |||      |||      :::
Db    930 AspLeuAsnAsnTyrValLeuTyrAlaThrGlyAsnAspGlyThrValLeuLysGlyGlu 949

Qy   1699 GCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGT 1758
                                  |||   |||:::      |||:::   |||   |||
Db    950 LysThrLeuGluProGlyArgTyrTyrLeuSerValTyrThrTyrAspAsnGlnSerGly 969

Qy   1759 GCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGC 1818
                      |||            ||||||  |||:::   :::      ::::::
Db    970 ---------ThrTyrThrValAsnValLysGlyAsnLeuLysAsnGluValLysGluThr 986

Qy   1819 GATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGC 1878
             :::   |||::::::|||
Db    987 AlaLysAspAlaIleLysGlu--------------------------------------- 993

Qy   1879 TATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGT 1938
                ||||||                           |||||||||
Db    994 ---ValGluAsn---------------------------AsnAsnAsp------------ 999

Qy   1939 AACTCCGGTAGCCCCGTATTCGATAAG------------------AACGGCCGTCTGATC 1980
                            |||||||||                  |||   ::::::::
Db   1000 ------------------PheAspLysAlaMetLysValAspSerAsnSerLysIleVal 1013

Qy   1981 GGT---CTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCC 2037
          |||   |||:::   |||            :::   |||   ||||||:::
Db   1014 GlyThrLeuSerAsnAspAspLeuLysAspIleTyrSerIleAspIleGlnAsnProSer 1033

Qy   2038 GATCTGCAGCGCACAATC------------AGCGTGGACATCCGCTACGTTCTCTTCATG 2085
          ||||||         :::            ::::::   :::   ::::::|||:::
Db   1034 AspLeuAsnIleValValGluAsnLeuAspAsnIleLysMetAsnTrpLeuLeuTyrSer 1053

Qy   2086 ATTGAC 2091
             |||
Db   1054 AlaAsp 1055

RESULT 11
NAGH_CLOPE
ID   NAGH_CLOPE     STANDARD;      PRT;  1628 AA.
AC   P26831;
DT   01-AUG-1992 (Rel. 23, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hyaluronoglucosaminidase precursor (EC 3.2.1.35) (Hyaluronidase)
DE   (Mu toxin).
GN   NAGH OR CPE0191.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CPN50;
RX   MEDLINE=94232189; PubMed=8177218;
RA   Canard B., Garnier T., Saint-Joanis B., Cole S.T.;
RT   "Molecular genetic analysis of the nagH gene encoding a hyaluronidase
RT   of Clostridium perfringens.";
RL   Mol. Gen. Genet. 243:215-224(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC   -!- FUNCTION: Putative virulence factor which is likely to act on
CC       connective tissue during gas gangrene.
CC   -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-linkages between N-
CC       acetyl-beta-D-glucosamine and D-glucuronate residues in
CC       hyaluronate.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- CAUTION: The partially purified protein from strain CPN50 is
CC       approximately 70 kDa smaller than the sequence indicated here.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M81878; AAA23259.1; -.
DR   EMBL; AP003185; BAB79897.1; -.
DR   PIR; S27540; S27540.
DR   InterPro; IPR002105; Dockerin_1.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00404; Dockerin_1; 2.
DR   Pfam; PF00754; F5_F8_type_C; 1.
DR   SMART; SM00231; FA58C; 1.
KW   Hydrolase; Glycosidase; Toxin; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1628       HYALURONOGLUCOSAMINIDASE.
FT   DOMAIN      703   1032       3 X APPROXIMATE REPEATS.
FT   REPEAT      703    748       1.
FT   REPEAT      848    894       2.
FT   REPEAT      987   1032       3.
FT   VARIANT     147    147       G -> A (IN STRAIN CPN50).
FT   VARIANT     172    175       KIQS -> EIKN (IN STRAIN CPN50).
FT   VARIANT     250    250       V -> M (IN STRAIN CPN50).
FT   VARIANT     548    548       A -> E (IN STRAIN CPN50).
FT   VARIANT     558    558       D -> E (IN STRAIN CPN50).
FT   VARIANT     614    614       G -> S (IN STRAIN CPN50).
FT   VARIANT     944    944       I -> V (IN STRAIN CPN50).
FT   VARIANT     950    950       N -> S (IN STRAIN CPN50).
FT   VARIANT     979    979       T -> I (IN STRAIN CPN50).
FT   VARIANT     982    982       I -> L (IN STRAIN CPN50).
FT   VARIANT    1042   1042       I -> F (IN STRAIN CPN50).
FT   VARIANT    1043   1628       MISSING (IN STRAIN CPN50).
SQ   SEQUENCE   1628 AA;  182474 MW;  D4252A2512BBED69 CRC64;

Alignment Scores:
Pred. No.:              2.75           Length:          1628
Score:                  121.00         Matches:         129
Percent Similarity:     32.24%         Conservative:    78
Best Local Similarity:  20.09%         Mismatches:      209
Query Match:            3.17%          Indels:          226
DB:                     1              Gaps:            36

US-10-008-355-1 (1-2139) x NAGH_CLOPE (1-1628)

Qy     526 ATCGTAGAGCCTTTCTATTCCAACAAC-----------GAATACTTCCTCATCGTCTAC 573
           :::|||:::  :::::::: |||             |||  |||         :::
Db     111 ValValAspAsnTyrPheAsnLysAsnIleProHisAspGluSerPhe---------Phe 127

Qy     574 GATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGA 633
            |||         |||    :::|||                 :::|||     |||
Db     128 AspGluLysMetAspAlaAsnIleValSerValLysAspGlyValIleGlyValIleGly 147

Qy     634 GGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTG 693
              |||||||||:::      :::            |||      :::::      |||
Db     148 GluAspThrAspSerAlaPheTyr-----------GlyValThrThrLeuLysHisVal 163

Qy     694 TATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCC 753
           :::               |||:::     |||    :::|||
Db     164 Phe-----------AsnGlnLeuGluGluGlyAsnLys-------------------- 172

Qy     754 GTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATC 813
                                 :::|||   ::::::|||||||||||||||     :::
Db     173 --------------------IleGlnSerPheArgAlaAspAspTyrAla---GluVal 184

Qy     814 GGTTTCCCGGGCAGTACGGATCGCTACCTC---ACTTCTTGGGGTGTGGAAGATCGTATC 870
                  |||      :::  |||             |||      |||||||||
Db     185 AlaHisArgGlyPheIleGluGlyTyrTyrGlyAsnProTrpSerAsnGluAspArgAla 204

Qy     871 GAA-----------------------------------------------AACGAGAAC 882
           |||                                                  ::::::
Db     205 GluLeuMetLysPheGlyGlyAspTyrLysLeuAsnGlnTyrValPheAlaProLysAsp 224

Qy     883 AATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCA 942
           :::|||                                         |||::::::

Db     225 AspPro-----------------TyrHisAsnSerLysTrpArgAspLeuTyrProGlu 238

Qy     943 GATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAG 1002
           ::::::   :::  ||||||           |||   ||||||       |||      |||
Db     239 GluLysLeuSerGluIleLys---------LysLeuAlaGlnValGlyAsnGluThrLys 255

Qy    1003 AATTCGATC-----------------GGTATGAACCGCGGTCTCGCT----------- 1032
           |||                          ||||||      :::
Db     256 AsnArgTyrValTyrAlaLeuHisProPheMetAsnAsnProValArgPheAspThrGlu 275

Qy    1033 --------------CGTCTTGACGTGATAGGTCGTAAG---------------CGTGCC 1062
                          |||   ||||||       |||
Db     276 GluAsnTyrGlnAsnAspLeuGlyValIleLysAlaLysPheThrGlnLeuLeuGluAsn 295

Qy    1063 GAGGAAAGAGCATTCGCA------------------------------------------ 1080
           :::   |||   ||||||
Db     296 AspValArgGlnPheAlaIleLeuAlaAspAspAlaSerAlaProAlaGlnGlyAlaSer 315

Qy    1081 -----------------------------GACTGGATCCGTAAGAACGGCAAGAGTGCT 1110
                                        |||:::   :::        :::   :::
Db     316 MetTyrValLysLeuLeuThrAspLeuThrArgTrpLeuGluGlu------GlnGlnSer 333

Qy    1111 GTCTATGGCGATGTATTGTCTTCTCTC-----------GAAAAGGCTTATAAGGAAGGA 1158
              |||   |||:::   :::  |||                       |||     |||
Db     334 ThrTyrProAspLeuLysThrAspLeuMetPheCysProSerAspTyrTyrGlyAsnGly 353

Qy    1159 GCCAAGGCC--------------AACCGTGAGATGACTTATTTGAGCGAGACGCTCTTC 1203
           :::    |||                 |||:::           :::|||       :::
Db     354 SerSerAlaGlnLeuLysGluLeuAsnLysAlaGluAspAsnValSerIleValMetThr 373

Qy    1204 GGTGGTACC---------GAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACA 1254
           ||||||             ||||||                 |||||||||
Db     374 GlyGlyArgIleTrpGlyGluValAspGlu------AsnPheAlaAsnAsnPheMetAsn 391

Qy    1255 AAT------CCTGATGCTCATGCCGGT------------------ATCCTCAAATCGCTT 1290
           |||             :::  |||   |||                  |||
Db     392 AsnIleSerThrGluGlyHisProGlyArgAlaProPhePheTrpIleAsnTrpProCys 411

Qy    1291 GACGACAAGTACAAAGACTACCTC-----------------------------CCCTCG 1320
              |||      |||   :::|||                              |||
Db     412 SerAspAsnSerLysGlnHisLeuIleMetGlyGlyAsnAspThrPheLeuHisProGly 431

Qy    1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
           :::|||          :::         ::::::                |||:::
Db     432 ValAspProSerLysIleAspGlyIleValLeuAsnProMetGlnGlnAlaGluAlaAsn 451

Qy    1381 AAG---------------CTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAA 1425
           |||                :::   |||        |||:::   |||   |||   :::
Db     452 LysSerAlaLeuPheAlaIleAlaAspTyrAlaTrpAsnIleTrpAspAsnLysGluGlu 471

Qy    1426 GGCGACACGAAG---------------AAGTATGCAGACTTC-----------GTATTC 1458
              |||                        ||||||   |||
Db     472 AlaAspGluAsnTrpAsnAspSerPheLysTyrMetAspHisGlyThrAlaGluGluThr 491

Qy    1459 GACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGAC--- 1515
           :::   |||:::                        |||   :::   ::::::||||||
Db     492 AsnSerSerLeuAlaLeuArgGluIleSerLysHisMetIleAsnGlnAsnMetAspGly 511

Qy    1516 ---AAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGC 1572
                 :::::::|||        |||    :::|||       |||
Db     512 ArgValArgProLeuGlnGluSerValGluLeuAlaProLysLeuGluAlaPheLysGln 531

Qy    1573 GTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATG---------GCCAATGCCTAT 1623
                 :::      :::|||:::     ||||||:::          |||
Db     532 LysTyrAspSerGlyAlaSerIleLysGluAspAlaLeuGluLeuIleAlaGluPheThr 551

Qy    1624 GCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGT 1683
              ::::::|||                    |||          :::      ||||||
Db     552 AsnLeuGlnLysAla---------------AlaAspTyrTyrLysAsnAsnProGlyAsn 566

Qy    1684 GCTCTGCCGAGCGATGCC------------------------AACTTCACCATGCGTATG 1719
                 |||                                           ||||||
Db     567 GluArgThrArgAspGlnIleIleTyrTrpLeuAsnCysTrpGluAspThrMetAspAla 586
```

```
Qy   1720 AGCTACGGCTCCATCAAGGGA--------------TATGAACCGCAGGACGGTGCCTGG 1764
               :::    |||   :::|||                       :::   ||||||
Db    587 AlaIleGlyTyrLeuLysSerAlaIleAlaIleGluGluGlyAspAspGluAlaAlaTrp 606

Qy   1765 TACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAG 1824
             ||||||   :::        :::|||      ||||||
Db    607 AlaAsnTyr---SerGluAlaGlnGlyAlaPheGluLys-------------------- 618

Qy   1825 TTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGT--------- 1875
                                                   :::|||   ||||||
Db    619 ---------------------------------SerLysThrTyrGlyPheHisTyr 626

Qy   1876 -----------CGCTATGCCGAGAACGGTCAGCTCCATATC----------------- 1905
                     |||||||||   |||      ||||||
Db    627 ValAspHisThrGluTyrAlaGluValGlyValGlnHisIleValProPheIleLysSer 646

Qy   1905 ------------------------------------------------------------ 1905

Db    647 MetGlyGlnAsnLeuSerValValIleGlySerIleValAspProAsnArgIleIleAla 666

Qy   1906 GCTTTCCTATCGAAC---AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGAT 1962
             ::::::||||||        |||         ||||||      :::    :::||||||
Db    667 ThrTyrIleSerAsnArgGlnAspAlaProThrGlyAsnProAspAsn---IlePheAsp 685

Qy   1963 AAGAAC 1968
             |||
Db    686 AsnAsn 687
```

```
RESULT 12
EFG_VIBCH
ID   EFG_VIBCH      STANDARD;      PRT;   698 AA.
AC   Q9KUZ7;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Elongation factor G (EF-G).
GN   FUSA OR VC0361.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- FUNCTION: This protein promotes the GTP-dependent translocation of
CC       the nascent protein chain from the A-site to the P-site of the
CC       ribosome.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC       EF-G/EF-2 SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE004124; AAF93534.1; -.
DR   HSSP; P13551; 1ELO.
DR   TIGR; VC0361; -.
DR   InterPro; IPR004540; EF-G.
DR   InterPro; IPR000640; EFG_C.
DR   InterPro; IPR004161; EFTU_D2.
DR   InterPro; IPR000795; EF_GTPbind.
DR   InterPro; IPR005225; Small_GTP.
DR   Pfam; PF00679; EFG_C; 1.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 1.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   TIGRFAMs; TIGR00484; EF-G; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   PROSITE; PS00301; EFACTOR_GTP; 1.
KW   Elongation factor; Protein biosynthesis; GTP-binding;
KW   Complete proteome.
FT   NP_BIND      17     24       GTP (BY SIMILARITY).
FT   NP_BIND      88     92       GTP (BY SIMILARITY).
FT   NP_BIND     142    145       GTP (BY SIMILARITY).
SQ   SEQUENCE   698 AA;  76927 MW;  F44FDA4DB8FC4ECC CRC64;

Alignment Scores:
Pred. No.:              3.3            Length:         698
Score:                  118.00         Matches:        120
Percent Similarity:     31.18%         Conservative:   83
Best Local Similarity:  18.43%         Mismatches:     202
Query Match:            3.09%          Indels:         246
DB:                     1              Gaps:           26

US-10-008-355-1 (1-2139) x EFG_VIBCH (1-698)
```

```
Qy     22 ATTCTTCTCGGAGCAGCCCTGCTGTTG---GGTGCTTCAGGGGTAGCCAAAGCCGACAAA 78
             :::|||   ||||||   ::::::         |||   |||||||||           :::
Db    105 ValLeuAspGlyAlaValValValPheCysGlyThrSerGlyValGluProGlnSerGlu 124

Qy     79 GGCATGTGG---------------------------------CTCCTCAACGAACTC 102
             :::|||                                   :::|||::::::
Db    125 ThrValTrpArgGlnAlaAspLysTyrGlyValProArgMetValPheValAsnLysMet 144

Qy    103 AATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTC 162
             ::::::            |||   :::|||   :::|||
Db    145 AspArgAlaGlyAlaAspPheLeuArgValValGly---------------------- 156

Qy    163 TACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATC----------------- 204
                                       ||||||    |||    |||
Db    157 ---GlnIleLysHisArgLeuGlyAlaAsnProValProIleGlnLeuAsnIleGlyAla 175

Qy    205 ---------TTCGGTGGCGGATGTACCGGTATCACAGTG-------------------- 234
                     |||   |||             |||   :::
Db    176 GluGluGluPheLysGlyValIleAspLeuIleLysMetLysAlaIleAsnTrpAsnGlu 195

Qy    235 TCCGATCAGGGCCTGATCTTTACC--------------------------------- 258
             :::|||||||||:::    ||||||
Db    196 AlaAspGlnGlyMetSerPheThrTyrGluGluIleProAlaAspMetLeuGluLeuAla 215

Qy    259 ---------AACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCAC 309
                      :::|||         |||   ::::::         :::
Db    216 GlnGluTrpArgAsnHisLeuValGluAlaAlaAlaGluAlaSerGluGluLeuMetGlu 235

Qy    310 GACTATCTGCGCGATGGTTTCGTTTCT----------------------------CGC 339
                ||||||   ||||||   :::|||                               |||
Db    236 LysTyrLeuGluAspGlyGluLeuSerGluValGluIleLysGlnAlaLeuArgGlnArg 255

Qy    340 ACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTG 399
             |||:::         |||:::   :::                  :::      |||   |||
Db    256 ThrIleAsnAsnGluIleValLeuAlaAlaCysGlySerAlaPheLysAsnLysGlyVal 275

Qy    400 AAG---GTAACGGACAAGGTA-------------------------GAAGGACAGCTC 429
             :::     |||   |||   |||                                   :::
Db    276 GlnAlaValLeuAspAlaValIleGluPheLeuProSerProThrAspValProAlaIle 295

Qy    430 AAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTG 489
             |||||||||   |||                                       :::|||
Db    296 LysGlyIleAspAsp---------------------------------ArgGluAsn 303
```

```
Qy     490 GCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAAC 549
           :::  :::    :::||||||:::|||                 |||||||||     :::
Db     304 SerValGluArgHisAlaAspAspAsn---------------GluProPhe------Ser 316

Qy     550 AACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCT 609
           :::         |||   |||       |||   |||
Db     317 SerLeuAlaPheLysIleAlaThrAspProPhe--------------------------- 327

Qy     610 CCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACG 669

Db     328 ---------------------------------------------------------Val 328

Qy     670 GGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGC 729
           |||       :::       |||||||||:::|||         |||
Db     329 GlySerLeuThrPheIleArgValTyrSerGlyValValAsnSerGlyAspAlaValTyr 348

Qy     730 AAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAG 789
                    |||      |||         :::         |||   |||         |||
Db     349 AsnSerValLysGlnLysLysGluArgPheGlyArgIleValGlnMetHisAlaAsnLys 368

Qy     790 GCTGAC-----------------------GACTATGCCATGACCATCGGT--------- 816
              |||                       |||   |||       ||||||
Db     369 ArgAspGluIleLysGluIleArgAlaGlyAspIleAlaAlaAlaIleGlyLeuLysAsp 388

Qy     816 ------------------------------------------------------------ 816

Db     389 ValThrThrGlyAspThrLeuCysAspProAsnHisValValIleLeuGluArgMetGlu 408

Qy     817 TTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAAC 876
           ||||||                                    |||       |||:::
Db     409 PhePro------------------------------------GluProValIleGlnIle 416

Qy     877 GAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATG 936
                    ||||||   :::                |||   ||||||            :::
Db     417 AlaValGluProArgSerLysAlaAspGlnGluLysMetGlyIleAlaLeuGlyLysLeu 436

Qy     937 AGCGCA---GATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGT------ 987
           :::|||   |||   :::   |||::::::      ::::::      |||:::
Db     437 AlaAlaGluAspProSerPheArgValGluThrAspAlaGluThrGlyGlnThrLeuIle 456

Qy     988 ------------------------------GCTAACTATTGGAAGAATTCGATCGGT 1014
                                         :::      |||         |||
Db     457 SerGlyMetGlyGluLeuHisLeuAspIleIleValAspArgMetLysArgGluPheGly 476

Qy    1015 ATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAG------CGTGCCGAGGAA 1068
           ::::::      :::   :::      |||      |||:::      |||   :::
Db     477 ValAspCysAsnValGlyLysProGlnValAlaTyrArgGluThrIleArgGlyLysSer 496

Qy    1069 AGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGC--------- 1119
                          ::::::|||::::::|||            ||||||
Db     497 GluValGluGlyLysPheValArgGlnSerGlyGlyArgGlyGlnTyrGlyHisValTrp 516

Qy    1119 ------------------------------------------------------------ 1119

Db     517 LeuLysIleGluProAlaGluProGlyGlnGlyPheValPheValAspAlaIleAlaGly 536

Qy    1120 -------------------------GATGTATTGTCTTCTCTCGAAAAGGCTTATAAG 1152
                                    |||         :::|||:::
Db     537 GlyValIleProLysGluPheIleAsnProValAlaLysGlyIleGluGluGlnMetAsn 556

Qy    1153 GAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACC 1212
              |||      |||         :::      :::      |||||||||   |||:::
Db     557 AsnGlyValLeuAlaGlyTyrProValLeuAspValLysAlaThrLeuPheAspGlySer 576

Qy    1213 ---GAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCAT 1269
              |||      ::::::   |||      :::|||   :::
Db     577 PheHisAspValAspSerSerGluMetAlaPheLysIleAlaGlySerMetAlaPheLys 596

Qy    1270 GCCGGTATCCTCAAATCGCTT--------------------------------------- 1290
              |||   |||::::::
Db     597 LysGlyAlaLeuGluAlaGlnProValLeuLeuGluProLeuMetLysValGluIleThr 616

Qy    1291 ------GACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAG------GTGCTG 1338
                 :::|||         |||   :::         |||:::|||:::       ::::::
Db     617 ThrProGluAspTrpMetGlyAspValValGlyAspLeuAsnArgArgArgGlyIleIle 636

Qy    1339 -----------------CCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
                             ||||||   |||   |||:::      :::::|||
Db     637 GluGlyMetAspGluGlyProAlaGlyLeuLysIleIleHisAlaLysValPro------ 654

Qy    1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAA--------------------- 1419
              |||   ::::::|||               |||   :::
Db     655 ---LeuSerGluMetPheGlyTyrAlaThrAspLeuArgSerAlaThrGlnGlyArgAla 673

Qy    1420 ---TTCAAAGGCGACACGAAGAAGTATGCAGAC 1449
              :::      :::      :::|||||||||
Db     674 SerTyrSerMetGluPheAlaGluTyrAlaAsp 684

RESULT 13
HLYA_ECOLI
ID   HLYA_ECOLI     STANDARD;      PRT;  1024 AA.
AC   P08715;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hemolysin, plasmid.
GN   HLYA.
OS   Escherichia coli.
OG   Plasmid IncI2 pHLY152.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hess J., Wels W., Vogel M., Goebel W.;
RT   "Nucleotide sequence of a plasmid-encoded hemolysin determinant and
RT   its comparison with a corresponding chromosomal hemolysin sequence.";
RL   FEMS Microbiol. Lett. 34:1-11(1986).
RN   [2]
RP   PALMITOYLATION OF LYS-564 AND LYS-690.
RX   MEDLINE=95099325; PubMed=7801126;
RA   Stanley P., Packman L.C., Koronakis V., Hughes C.;
RT   "Fatty acylation of two internal lysine residues required for the
RT   toxic activity of Escherichia coli hemolysin.";
RL   Science 266:1992-1996(1994).
RN   [3]
RP   PALMITOYLATION OF LYS-564 AND LYS-690.
RX   MEDLINE=96404790; PubMed=8808931;
RA   Ludwig A., Garcia F., Bauer S., Jarchau T., Benz R., Hoppe J.,
RA   Goebel W.;
RT   "Analysis of the in vivo activation of hemolysin (HlyA) from
RT   Escherichia coli.";
RL   J. Bacteriol. 178:5422-5430(1996).
CC   -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC       CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC       DEFINED.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CC       CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
CC       ACTIVITY.
CC   -!- DOMAIN: THE THREE TRANSMEMBRANE DOMAINS ARE BELIEVED TO BE
CC       INVOLVED IN PORE FORMATION BY THE CYTOTOXIN.
CC   -!- PTM: PALMITOYLATED BY HLYC. THE TOXIN ONLY BECOMES ACTIVE WHEN
CC       MODIFIED.
CC   -!- DISEASE: THE HEMOLYSIN OF E.COLI IS PRODUCED PREDOMINANTLY BY
CC       STRAINS CAUSING EXTRAINTESTINAL INFECTIONS, SUCH AS THOSE OF THE
CC       URINARY TRACT.
CC   -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   -----------------------------------------------------------------------
DR   EMBL; M14107; AAA98233.1; -.
DR   InterPro; IPR001343; Hemlysn_Ca_bind.
DR   InterPro; IPR003355; RTXtoxin_N.
DR   InterPro; IPR003995; RtxA.
DR   Pfam; PF00353; hemolysinCabind; 6.
DR   Pfam; PF02382; RTX; 1.
DR   PRINTS; PR00313; CABNDNGRPT.
DR   PRINTS; PR01488; RTXTOXINA.
DR   PROSITE; PS00330; HEMOLYSIN_CALCIUM; 4.
KW   Hemolysis; Toxin; Cytolysis; Cytotoxin; Repeat; Calcium;
KW   Transmembrane; Lipoprotein; Palmitate; Plasmid.
FT   TRANSMEM    238    260       POTENTIAL.
FT   TRANSMEM    268    327       POTENTIAL.
FT   TRANSMEM    365    411       POTENTIAL.
FT   DOMAIN      724    870       16 X REPEATS, GLY-RICH.
FT   REPEAT      724    729       1.
FT   REPEAT      733    738       2.
FT   REPEAT      742    747       3.
FT   REPEAT      751    756       4.
FT   REPEAT      760    765       5.
FT   REPEAT      769    774       6.
FT   REPEAT      778    783       7.
FT   REPEAT      787    792       8.
FT   REPEAT      796    801       9.
FT   REPEAT      807    812       10.
FT   REPEAT      817    822       11.
FT   REPEAT      826    831       12.
FT   REPEAT      835    840       13.
FT   REPEAT      844    849       14.
FT   REPEAT      856    861       15.
FT   REPEAT      865    870       16.
FT   LIPID       564    564       PALMITATE.
FT   LIPID       690    690       PALMITATE.
SQ   SEQUENCE   1024 AA;  110201 MW;  83944917F76C945B CRC64;

Alignment Scores:
Pred. No.:                 4.04          Length:        1024
Score:                     117.50        Matches:       125
Percent Similarity:        33.87%        Conservative:  86
Best Local Similarity:     20.06%        Mismatches:    246
Query Match:               3.08%         Indels:        166
DB:                        1             Gaps:          27

US-10-008-355-1 (1-2139) x HLYA_ECOLI (1-1024)

Qy     397 GTGAAGGTAACGGACAAGGTAGAAGGACAGCTC---AAGGGTATCACTGACGAGATGGAG 453
           |||:::::|||    ||||||   |||    :::   |||||||||:::        :::
Db     273 ValGluLeuThrThrLysValLeuGlyAsnValGlyLysGlyIleSerGlnTyrIleIle 292

Qy     454 CGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAG 513
                    |||   ||||||    :::               |||
Db     293 AlaGlnArgAlaAlaGlnGlyLeuSerThrSerAlaAlaAlaAlaGlyLeuIleAlaSer 312

Qy     514 AACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTAC 573
                                :::   |||                     ||||||   :::
Db     313 AlaValThrLeuAlaIleSerProLeu---------------SerPheLeuSerIleAla 327

Qy     574 GATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGA 633
           |||   ||||||              :::                    |||   |||
Db     328 AspLysPheLysArgAlaAsnLysIleGluGluTyrSerGlnArgPheLysLysLeuGly 347

Qy     634 GGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTG 693
              |||  |||:::   :::      |||  :::   |||   |||      |||
Db     348 TyrAspGlyAspSerLeuLeuAlaAlaPheHis-LysGluThrGlyAlaIleAspAlaSe 367

Qy     694 TATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCC 753
            :::          :::|||          :::   :::     |||:::       :::
Db     367 rLeuThrThrIleSerThrValLeuAlaSerValSerSerGlyIleSerAlaAlaAla-- 386

Qy     754 GTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATC 813
                                            |||   |||         |||      ::
Db     387 ---------------------------ThrThrSerLeuValGlyAlaProValSerAl 397

Qy     814 GGTTTCCCGGGCAGTA------------------------------------------- 829
           ::::    ||||||
Db     397 aLeuValGlyAlaValThrGlyIleIleSerGlyIleLeuGluAlaSerLysGlnAlaMe 417

Qy     830 -------CGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAAC 882
                  :::|||:::   :::          |||         :::   |||:::
Db     417 tPheGluHisValAlaSerLysMetAlaAspValIle-AlaGluTrpGluLysLysHisG 437

Qy     883 AATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCAT---------CTGGAAGGAAGCC 933
              :::|||     :::   |||   |||||||||         ||||||   :::
Db     437 lyLysAsnTyrPheGluAsnGlyTyrAspAlaArgHisAlaAlaPheLeuGluAspAsnP 457

Qy     934 ATGAGCGCAGATCAGGCTAC------CCGTATCAAATATGCCAGCAAGTATGCTCAGAGT 987
             :::      |||   |||             |||   :::   :::   |||      :::
Db     457 heLysIleLeuSerGlnTyrAsnLysGluTyrSerValGluArgSerValLeuIle-Thr 476

Qy     988 GCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATA 1047
              :::|||         ||||||   .   |||:::   |||
Db     477 GlnGlnHisTrpAspThrLeuIleGlyGluLeuAlaGlyValThrArg---------Asn 493

Qy    1048 GGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGT 1107
           |||   |||               :::::::::   |||:::
Db     494 GlyAspLysThrLeuSerGlyLysSerTyrIleAspTyr-------------------- 506

Qy    1108 GCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCC 1167
                                                    |||:::||||||   :::
Db     507 -------------------------------------TyrGluGluGlyLysArgLeu 513

Qy    1168 AACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTT 1227
           ::::::                          :::   :::|||
Db     514 GluLysLys***AspGluPheGlnLysGlnValPhe----------------------- 525

Qy    1228 GCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCG 1287
                                                  |||         |||   :::
Db     526 --------------------------------AspProLeuLysGlyAsnIleAspLeu 534

Qy    1288 CTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATG 1347
              |||   |||            |||      :::         :::   |||
Db     535 SerAspSerLysSerSerThrLeuLeuLysPheValThrProLeuLeuThrProGly--- 553

Qy    1348 CTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTA 1407
              :::   :::|||   |||      :::   |||            ::::::   :::
Db     554 ---GluGluIleArgGluArgArgGlnSerGlyLys-----------TyrGluTyrIle 568

Qy    1408 ATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAG--- 1464
              :::            ||||||      |||                  |||   ||||||
Db     569 ThrGluLeuLeuValLysGlyValAspLysTrpThrValLysGlyValGlnAspLysGly 588

Qy    1465 AGTGTGGTTCCTTATAGCGACAAGTTC---CATGCCATGCTCAAGTCCATGGACAAGGAA 1521
           :::|||      ||||||:::         ||||||         |||:::
Db     589 AlaValTyrAspTyrSerAsnLeuIleGlnHisAla---------SerValGlyAsnAsn 605

Qy    1522 AAGTTTGCCAAG---GCTATCGAGAAA--------------GATCCGGCAGTAGAGCTT 1563
           ::::::   :::         ||||||                  |||      |||   |||
Db     606 GlnTyrArgGluIleArgIleGluSerHisLeuGlyAspGlyAspAspLysValPheLeu 625

Qy    1564 TCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTAT 1623
           |||                                          ||||||   |||
Db     626 SerAla-----------------------------------GlySerAlaAsnIleTyr 633

Qy    1624 GCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAG-----------ATG 1671
           |||      ||||||      :::   :::              . :::
Db     634 Ala---GlyLysGlyHisAspValValTyrTyrAspLysThrAspThrGlyTyrLeuThr 652

Qy    1672 TACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATG---CGTATGAGCTACGGC 1728
              |||                           |||:::|||:::   |||:::      |||
Db     653 IleAspGlyThrLysAlaThrGluAlaGlyAsnTyrThrValThrArgValLeuGlyGly 672

Qy    1729 TCCATCAAG--------------------------------------GGATATGAACCG 1749
              :::|||                                        |||
Db     673 AspValLysValLeuGlnGluValValLysGluGlnGluValSerValGlyLysArgThr 692
```

```
Qy   1750 CAGGACGGTGCCTGGTACAACTAT---------CATACGACAGGCAAGGGCGTATTGGAG 1800
            :::         :::   :::|||             |||      ||||||   :::    |||
Db    693 GluLysThrGlnTyrArgSerTyrGluPheThrHisIleAsnGlyLysAsnLeuThrGlu 712

Qy   1801 AAGCAGGAT-----------------------------CCTAAGAGCGATGAGTTTGCC 1830
               :::                                  ::::::|||:::|||
Db    713 ThrAspAsnLeuTyrSerValGluGluLeuIleGlyThrThrArgAlaAspLysPhe--- 731

Qy   1831 GTACAGGAGAATATCCTCGACCTCTTC--------------------------------- 1857
                           |||:::|||
Db    732 PheGlySerLysPheThrAspIlePheHisGlyAlaAspGlyAspAspLeuIleGluGly 751

Qy   1858 CGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCG 1917
                |||         |||    :::     |||
Db    752 AsnAspGlyAsnAspArgLeuTyrGlyAspLysGly------------------------ 763

Qy   1918 AACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTC------GATAAGAACGGC 1971
          |||:::   ::::::|||||||||                 ::::::      |||    |||
Db    764 AsnAspThrLeuSerGlyGlyAsnGlyAspAspGlnLeuTyrGlyGlyAspGlyAsnAsp 783

Qy   1972 CGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTC 2031
          :::||||||||||:::|||    :::                      ||||||   ||||||
Db    784 LysLeuIleGlyValAlaGlyAsnAsnTyrLeuAsnGlyGlyAspGlyAspAspGluPhe 803

Qy   2032 GAA 2034
          :::
Db    804 Gln 804
```

RESULT 14
TEA1_SCHPO

```
ID   TEA1_SCHPO     STANDARD;      PRT;  1147 AA.
AC   P87061;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tip elongation aberrant protein 1 (Cell polarity protein teal).
GN   TEA1 OR SPCC1223.06.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97344085; PubMed=9200612;
RA   Mata J., Nurse P.;
RT   "teal and the microtubular cytoskeleton are important for generating
RT   global spatial order within the fission yeast cell.";
RL   Cell 89:939-949(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: CELL POLARITY PROTEIN. MAY ACT AS AN END MARKER,
CC       DIRECTING THE GROWTH MACHINERY TO THE CELL POLES. MAY ALSO
CC       INFLUENCE MICROTUBULAR ORGANIZATION, AFFECTING THE MAINTENANCE OF
CC       A SINGLE CENTRAL AXIS.
CC   -!- SUBCELLULAR LOCATION: PRESENT AT BOTH POLES OF THE CELL THOUGHOUT
CC       THE CELL CYCLE WHETHER THEY ARE GROWING OR NOT. LOCATED AT THE
CC       ENDS OF MICROTUBULES GROWING TOWARDS THE CELL POLES. AN INTACT
CC       MICROTUBULAR SKELETON IS REQUIRED TO MAINTAIN THE LOCATION AT THE
CC       CELL TIPS; ON THE OTHER HAND, A NORMALLY ORGANIZED ACTIN
CC       CYTOSKELETON IS NOT REQUIRED.
CC   -!- SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y12709; CAA73246.1; -.
DR   EMBL; AL031579; CAA20875.1; -.
DR   InterPro; IPR001798; Kelch.
DR   Pfam; PF01344; Kelch; 5.
KW   Repeat; Microtubules; Coiled coil.
FT   REPEAT       94    144       KELCH 1.
FT   REPEAT      146    198       KELCH 2.
FT   REPEAT      254    303       KELCH 3.
FT   REPEAT      305    351       KELCH 4.
FT   REPEAT      355    402       KELCH 5.
FT   DOMAIN      611    649       COILED COIL (POTENTIAL).
FT   DOMAIN      716    838       COILED COIL (POTENTIAL).
FT   DOMAIN      879    990       COILED COIL (POTENTIAL).
FT   DOMAIN     1084   1105       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1147 AA;  127436 MW;  7BE65F6C666EF4F8 CRC64;
```

Alignment Scores:

Pred. No.:	4.53	Length:	1147
Score:	117.00	Matches:	92
Percent Similarity:	36.99%	Conservative:	70
Best Local Similarity:	21.00%	Mismatches:	168
Query Match:	3.06%	Indels:	109
DB:	1	Gaps:	16

US-10-008-355-1 (1-2139) x TEA1_SCHPO (1-1147)

```
Qy    473 AGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAG 532
          |||   :::   :::   |||                :::   ||||||:::||||||:::
Db    460 ArgAlaSerAsnAspLeuProSerProValValProThrArgSerAsnSerSer------ 477

Qy    533 AGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACG------------------ 574
          |||:::      |||:::   |||   :::|||   |||:::
Db    478 SerIleLeuGlnProSerTyrAsnLeuAsnSerHisSerSerAspArgArgAsnThrAsn 497

Qy    575 ---------------------------------ATGTATTCAAGGACGTTCGTATGGTAT 601
                                            :::   ::::::
Db    498 AspAspAspGlnSerSerLeuAsnSerGlnGlnLeuSerAsnGlnAlaLysAlaGlnGly 517

Qy    602 TTGCTCCTCCCAGCTCTG---TAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGC 658
             :::   |||   |||      |||   |||                  |||   |||
Db    518 GluValSerProThrLeuSerPheValProSerSerHisSerMetGluGlnGlyAsnGly 537

Qy    659 CGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGG 718
             |||         |||      :::|||   |||
Db    538 SerValAlaSerAlaAsnAsnAlaGlnSerGluAla------------------------ 549

Qy    719 CCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGC 778
```

```
                        |||:::||||||  ||||||       :::   :::   :::|||
Db     550 ------AlaThrArgSerIleAsnSerIleSerGluValSerGluValArgPheProGlu 567

Qy     779 AAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCT 838
           :::                            |||  :::  |||
Db     568 Gln-----------------------------SerSerValLysThrValAspGluArg 577

Qy     839 ACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAG 898
              ||||||  |||       :::|||  |||  |||         ||||||  |||
Db     578 LysSerLeuAspGly------ArgIleThrSerValThrLeuGluThrLeuValGluLys 595

Qy     899 TTCGCGGTATCAAGCAAGGCATC----------TGGAAGGAAGCC------------ATG 936
           ::::::    ||||||               |||  ::::::              :::
Db     596 TyrSerGluLeuSerLysGlnGlnIleValGluTrpPheLysSerLysLeuTyrGluIle 615

Qy     937 AGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTAT 996
              |||  |||:::::|||         ::::::     :::  ||||||
Db     616 LeuArgAspSerAlaSerLysIleAspSerLeuThrGluLysLeuLysValAlaAsnAla 635

Qy     997 TGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAG 1056
              ||||||:::                |||  ::::::  :::
Db     636 GluLysAsnAlaAlaLeuCysGluAlaAlaLeuGluLysValProLeuAlaLysHisAsn 655

Qy    1057 CGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCT------ 1110
           :::      :::     |||:::          ::::::|||   :::|||
Db     656 LysLeuSerAspGlyThrPheSerThrProAspLysGluAsnValGlnSerThrAsnAsp 675

Qy    1111 ---------------------------------------------GTCTATGGCGATGTA 1125
                                                       |||      :::
Db     676 AlaHisIleMetGlnGluAsnPheSerLeuHisLysAlaLeuGluValMetArgGluThr 695

Qy    1126 TTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTAT 1185
              |||  |||:::|||     |||:::             :::  :::|||:::
Db     696 SerSerAspLeuAspLysGlnLeuLysAspAlaThrAlaSerGlnLysGluLeu------ 713

Qy    1186 TTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCA 1245
                                         :::|||:::   :::   |||
Db     714 ------------------------------IleValGlnThrSerSerPheGlnLysGlu 723

Qy    1246 TTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAA 1305
           |||         :::  |||    |||   |||   |||   |||   :::.
Db     724 LeuValGluGluArgGluArgHisAsnAlaIleSerLysArgLeuGlnGlu--------- 740

Qy    1306 GACTACCTCCCCTCGCTC------GACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTA 1359
                 :::  ||||||     ||||||::::::|||         |||:::
Db     741 ------IleGluSerLeuTyrArgAspArgGluLeuLeuValThrAsnLeuGlu------ 756

Qy    1360 CGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAA 1419
                               |||:::|||  |||              :::
Db     757 -----------------AspGlnLeuValAspGlnThrValThrIleAsnLysPheAla 770

Qy    1420 TTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTAT 1479
           |||:::   |||          :::           |||:::   ::::::
Db     771 PheGluArgAspGlnPheArgGluArgSerMetGlyPheGluAsnThrIleLysAspLeu 790

Qy    1480 AGCGACAAG-----------------------------TTCCATGCCATGCTCAAGTCC 1509
           :::   |||                             .  |||      |||:::|||
Db     791 ThrArgLysMetGluAlaThrAspMetLeuAsnValSerLeuHisGluSerLeuArgSer 810

Qy    1510 ATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAG 1569
           :::      |||             |||            :::   |||   |||
Db     811 ValGlnThrGluAsnSerGluLeuValThrGlu------------MetAlaLeuLeuLys 826

Qy    1570 AGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTAT 1623
           :::   :::         :::||||||      ||||||  ||||||   |||
Db     827 AlaGluLeuValLysLysGlnAlaIle---IleAspAlaAsnAlaAsnIleTyr 843
```

RESULT 15
YK82_YEAST

```
ID   YK82_YEAST     STANDARD;      PRT;   1169 AA.
AC   P36170;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   Hypothetical 122.2 kDa protein in SIR1 3'region precursor.
GN   YKR102W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Gaillon L., Dujon B.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: CELL WALL (BY SIMILARITY).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY).
CC   -!- SIMILARITY: STRONG, TO YEAST PROTEIN FLO1.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z28327; CAA82182.1; -.
DR   PIR; S38181; S38181.
DR   SGD; S0001810; YKR102W.
KW   Hypothetical protein; Glycoprotein; Repeat; Cell wall; Signal.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27   1169       HYPOTHETICAL PROTEIN YKR102W.
FT   DOMAIN      274    609       SER/THR-RICH.
FT   CARBOHYD    122    122       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    157    157       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    389    389       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    452    452       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    656    656       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    686    686       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    879    879       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1092   1092       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1099   1099       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1169 AA;  122164 MW;  940C6DFB3569C669 CRC64;
```

Alignment Scores:

Pred. No.:	4.92	Length:	1169
Score:	116.50	Matches:	161
Percent Similarity:	33.56%	Conservative:	87
Best Local Similarity:	21.79%	Mismatches:	287
Query Match:	3.05%	Indels:	204
DB:	1	Gaps:	31

US-10-008-355-1 (1-2139) x YK82_YEAST (1-1169)

```
Qy      89 TCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCTCC 148
            ||||||   :::|||        :::      ||||||      :::
Db     392 SerSerPheSerSerSerSerGluValCysThrGluCysThrGluThr------------ 407

Qy     149 CGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCA------TTGCCAATGCCGTGGTTA 202
                       :::  |||||||||   |||                   |||
Db     408 ------------GluSerThrSerThrSerThrProTyrValThrSerSerSerTrpSer 423

Qy     203 TCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACC 262
            ||||||            :::::|||                        |||
Db     424 SerSerGluVal---------CysThrGluCys-----------------------Thr 432

Qy     263 ACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCG 322
              |||      |||     |||      |||         |||       :::|||
Db     433 GluThrGluSerThr-SerTyrValThrProTyrValSerSerSerThrAlaAlaAlaAs 452

Qy     323 ATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGT 382
            :::                                  |||      |||     |||||
Db     452 nTyr--------------------------ThrSerSerPheSerSerSerSerGluVa 463
```

```
Qy     383 ATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTG 442
           |         :::       |||          :::          |||:::      |||
Db     463 lCysThrGluCysThrGlu------ThrGluSerThrSerThrSerThrProTyr----- 479

Qy     443 ACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAA 502
                          :::|||:::  |||||| |||||||||  |||      :::   |||:
Db     480 -ValThrSerSerSerSerSerSer-GluValCysThrGluCysThrGluThrGluS 499

Qy     503 AT----------------------------------------GCAGACGAGAACCAAC 520
           ::                                          |||:::    |||:::
Db     499 erThrSerTyrValThrProTyrValSerSerSerThrAlaAlaAlaAsn-TyrThrSer 518

Qy     521 TCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTAT 580
           |||   |||   |||        :::   |||       |||    :::    ||||||
Db     519 SerPheSerSerSerSerGluValCysThrGluCysThrGluThrGluSerThrSerThr 538

Qy     581 TCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATA 640
           |||       :::|||                  :::      |||    :::|||
Db     539 SerThrProTyrValThrSerSerSerTrpSerSerSerGluValCysThrGluCysThr 558

Qy     641 CGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCG 700
              |||          ||||||         |||:::    :::|||
Db     559 GluThrGluSerThrSerTyrValThrProTyrValSerSerSerThrAla--------- 575

Qy     701 GTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACT 760
                                   |||       :::      |||         :::
Db     576 -----------------------AlaAlaAsnTyrThrSerSerPheSerSerSerSer 587

Qy     761 TCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCC 820
              :::         |||      |||      |||:::        |||        :::
Db     588 GluValCysThrGluCysThrGluThrGluSerThrSerThrSerThrProTyrAlaThr 607

Qy     821 CGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGA 880
              :::         ||||||||                    |||    |||
Db     608 SerSerThrGlyThrAlaThrSerPheThrAla---------SerThrSerAsnThrMet 624

Qy     881 ACAATCCTCGTATCG--------------AAGTTCGCGGTATCAAGCAAGGCATCTGGA 925
           |||   ||||||                 |||:::::::||||||       |||
Db     625 ThrSerLeuValGlnThrAspThrThrValSerPheSerLeuSerSerThrValSer--- 643

Qy     926 AGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGA 985
                                  |||   |||::::::   ::::::
Db     644 -----------GluHisThrAsnAlaProThrSerSerValGluSerAsnAlaSerThr 659

Qy     986 GTGCTAACTATT------GGAAGAATTCGATCGGTATGAACCGCGGTC----------- 1027
              ::::::          |||   ::::::|||      |||:::
Db     660 PheIleSerSerAsnLysGlySerValLysSerTyrValThrSerSerIleHisSerIle 679

Qy    1028 --------------------TCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGG 1066
                               |||   :::      |||||||||   |||
Db     680 ThrProMetTyrProSerAsnGlnThrValThrSerSerSerValValSerThrProIle 699

Qy    1067 AAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTAT 1126
                   |||         |||                  :::    |||
Db     700 ThrSerGluSerSerGluSerSerAlaSerValThrIleLeuProSer------------ 715

Qy    1127 TGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATT 1186
                                 |||       |||    :::|||:::
Db     716 -----------------ThrIleThrSerGluPheLysProSer--------------- 724

Qy    1187 TGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCAT 1246
                  ::::::   ||||||                   :::   |||   ||||||:::
Db     725 ---ThrMetLysThrLysValValSer------------IleSerSerSerProThrAsn 739

Qy    1247 TGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAG 1306
                                                  |||   ||||||
Db     740 ------------------------------------------LeuIleThrSerTyrAsp 745

Qy    1307 ACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGC 1366
           |||||||||         |||||||||                        |||    :::
Db     746 ThrThrSerLysAspSerThrValGlySerSerThrSerSerValSerLeuIleSerSer 765

Qy    1367 GTATCCCTGCCGACAAGC----------------------------TCCCCGATATAT 1396
           :::||||||||||:::|||                          |||   |||
Db     766 IleSerLeuProSerSerTyrSerAlaSerSerGluGlnIlePheHisSerSerIleVal 785

Qy    1397 TCA-----------------------------------AGAATGTAATCGACAAGAAAT 1420
           |||                                    ::::::   |||:::   :::
Db     786 SerSerAsnGlyGlnAlaLeuThrSerPheSerSerThrLysValSerSerSerGluSer 805

Qy    1421 TCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTC------ 1474
           |||::::::   |||   |||       ||||||      :::   :::
Db     806 SerGluSerHisArgThrSerProThrThrSerSerGluSerGlyIleLysSerSerGly 825

Qy    1475 ------------CTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAA 1522
                          :::|||||||||                        |||
Db     826 ValGluIleGluSerThrSerThrSerSerPheSerPheHisGluThrSerThrAlaSer 845

Qy    1523 AGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTG 1582
           :::        ::::::|||   :::                   |||   |||
Db     846 ThrSerValGlnIleSerSerGlnPheValThrProSerSerPro-------------- 860

Qy    1583 CTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGC 1642
              ::::::          |||                            |||
Db     861 ---IleSerThrValAlaPro--------------------------ArgSerThrGly 870

Qy    1643 GTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTC-------------- 1687
           :::   |||                  :::   :::   :::
Db     871 LeuAsnSerGlnThrGluSerThrAsnSerSerLysGluThrMetSerSerGluAsnSer 890

Qy    1688 TGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAAC 1747
                    ||||||:::|||               ||||||:::|||
Db     891 AlaSerValMetProSerSerSer---------AlaThrSerProLys------ThrGly 905

Qy    1748 CGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGG 1807
           :::   |||         |||:::                              |||
Db     906 LysValThrSerAspGluThrSer-----------------------------SerGly 915

Qy    1808 ATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAA 1867
                   |||       :::      ||||||      |||   |||   |||      :::
Db     916 PheSerArgAspArgThrThrValTyrArgMetThrSerGluThrProSerThrAsnGlu 935

Qy    1868 ------ACTATGGTCGCTATG-----------------------CCGAGAACGGTCAGC 1897
                 |||::::::   :::                             ||||||||
Db     936 GlnThrThrLeuIleThrValSerSerCysGluSerAsnSerCysSerAsnThrValSer 955

Qy    1898 TCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACT-------------- 1942
           |||      :::|||         |||||||||         :::|||
Db     956 SerAlaValValSerThrAlaThrThrThrIleAsnGlyIleThrThrGluTyrThrThr 975

Qy    1943 ------CCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATG 1996
                 |||::::::            :::   |||      |||   |||   |||
Db     976 TrpCysProLeuSerAlaThrGluLeuThrThrVal------SerLysLeuGluSerGlu 993

Qy    1997 GCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGC 2047
                    |||      |||||||||      :::     :::||||||
Db     994 GluLysThrThrLeuIleThrValThrSerCysGluSerGlyValCysSer 1010
```

Search completed: May 16, 2003, 12:29:53
Job time : 88.5 secs

```
                    GenCore version 5.1.4_p5_4578
           Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:          May 16, 2003, 12:21:16 ; Search time 164.5 Seconds
                                           (without alignments)
                                           5358.478 Million cell updates/sec

Title:           US-10-008-355-1
Perfect score:   3820
Sequence:        1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters:      1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110338_9352/app_query.fasta_1.2311
-DB=SPTREMBL_21 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_238_@runat_16052003_110338_9352 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       SPTREMBL_21:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	907.5	23.8	716	16	Q9PC94	Q9pc94 xylella fas
	2	206	5.4	1349	4	Q8WWQ4	Q8wwq4 homo sapien
	3	177	4.6	800	3	Q8TFG4	Q8tfg4 schizosacch
	4	157	4.1	957	4	O14651	O14651 homo sapien
	5	157	4.1	1217	4	Q9UKW9	Q9ukw9 homo sapien
	6	156.5	4.1	1029	4	Q8TDH7	Q8tdh7 homo sapien
	7	156	4.1	1275	4	Q9UQ36	Q9uq36 homo sapien
	8	156	4.1	1783	4	O15038	O15038 homo sapien
	9	156	4.1	1791	4	O60382	O60382 homo sapien
	10	156	4.1	2752	4	Q9UQ35	Q9uq35 homo sapien
	11	154.5	4.0	1236	3	Q9C105	Q9c105 schizosacch
	12	151	4.0	2448	4	Q8WWQ5	Q8wwq5 homo sapien
	13	149	3.9	528	6	Q29071	Q29071 sus scrofa
	14	147	3.8	1037	3	Q8WZX1	Q8wzx1 neurospora
	15	146.5	3.8	1368	16	Q99ZW2	Q99zw2 streptococc
	16	141.5	3.7	1795	5	O76894	O76894 drosophila
	17	141	3.7	948	3	O74346	O74346 schizosacch
	18	139.5	3.7	513	4	O43418	O43418 homo sapien
	19	139.5	3.7	1802	3	Q04051	Q04051 saccharomyc
c	20	138	3.5	684	5	Q9VTT7	Q9vtt7 drosophila
	21	137	3.6	3570	4	Q99552	Q99552 homo sapien
c	22	136	3.4	744	3	Q8TFG9	Q8tfg9 schizosacch
	23	135.5	3.5	589	4	O43419	O43419 homo sapien
	24	135.5	3.5	2296	4	Q9UHA8	Q9uha8 homo sapien
c	25	135	3.4	521	17	Q96Z01	Q96z01 sulfolobus
	26	134	3.5	477	4	Q14887	Q14887 homo sapien
	27	133.5	3.5	1104	2	Q06812	Q06812 micrococcus
	28	131.5	3.4	901	4	Q9H195	Q9h195 homo sapien
	29	131.5	3.4	991	17	Q8ZVU0	Q8zvu0 pyrobaculum
	30	131	3.4	648	4	O14760	O14760 homo sapien
	31	131	3.4	744	3	Q8TFG9	Q8tfg9 schizosacch
	32	130	3.4	1175	16	Q8RHT3	Q8rht3 fusobacteri
	33	129.5	3.4	1182	5	Q9U7F0	Q9u7f0 spironucleu
	34	127	3.3	2232	5	P91365	P91365 caenorhabdi
	35	126.5	3.3	666	17	O27798	O27798 methanobact
	36	126.5	3.3	3178	5	Q969D4	Q969d4 caenorhabdi
	37	126	3.3	953	2	Q48502	Q48502 lactobacill
	38	125	3.3	1440	5	Q9TYU6	Q9tyu6 caenorhabdi
	39	125	3.3	2055	5	Q8T5C7	Q8t5c7 plasmodium
	40	124.5	3.3	735	16	Q8U8L4	Q8u8l4 agrobacteri
	41	124.5	3.3	799	5	O18119	O18119 caenorhabdi
	42	124	3.2	1325	5	Q9BKV7	Q9bkv7 leishmania
	43	124	3.2	1808	10	Q9LQJ8	Q9lqj8 arabidopsis
	44	123.5	3.2	713	2	O85051	O85051 moraxella c
	45	123	3.2	1289	16	Q9X087	Q9x087 thermotoga

ALIGNMENTS

```
RESULT 1
Q9PC94
ID   Q9PC94      PRELIMINARY;      PRT;   716 AA.
AC   Q9PC94;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Xf1887.
GN   XF1887.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
```

```
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004008; AAF84693.1; -.
DR   MEROPS; S46.001; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   716 AA;  79375 MW;  EAF086E2315BBDFC CRC64;

Alignment Scores:
Pred. No.:               1.02e-54     Length:         716
Score:                   907.50       Matches:        239
Percent Similarity:      49.21%       Conservative:   135
Best Local Similarity:   31.45%       Mismatches:     291
Query Match:             23.76%       Indels:         95
DB:                      16           Gaps:           21

US-10-008-355-1 (1-2139) x Q9PC94 (1-716)

Qy       1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
             |||:::      |||   |||:::    :::   |||   |||:::           |||
Db       1 MetArgPheAsnLeuLeuSerLeuSerValLeuAlaThrLeuIleThrValAspSer--- 19

Qy      61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
                 |||   :::|||||||||:::         :::|||      |||
Db      20 ---ThrHisAlaGlyGluGlyMetTrpValProGlnGlnLeu---ProGluIleAlaGly 37

Qy     121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
              :::::::::      |||         |||         :::   |||   :::
Db      38 ProLeuLysGlnAlaGlyLeuGlnLeuSerProGluGlnLeuSerAsnLeuThrGlyAsp 57

Qy     181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
              :::         |||   :::         |||   ||||||         ||||||
Db      58 ProMetGlyAlaValValSerLeu------GlyAsnCysThrAlaSerLeuValSerPro 75

Qy     241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
           :::||||||:::   |||||||||||||||   |||||||||||||||         ||||||
Db      76 GluGlyLeuValIleThrAsnHisHisCysAlaTyrGlyAlaIleGlnLeuAsnSerThr 95

Qy     301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
              :::   :::::::::||||||   :::      |||         :::|||:::
Db      96 ProLysLysAsnLeuIleLysGluGlyPheAsnAlaLeuThrGlnAlaAspGluValSer 115

Qy     361 ATT---CCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTA 417
                 |||         :::      |||   :::|||         ||||||   :::
Db     116 AlaGlyProAsnAlaArgIleTyrValLeuGluGlnIleThrAspValThrAlaGlnAla 135

Qy     418 GAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTA 477
           :::      |||            ::::::      :::         |||   |||
Db     136 LysAlaAlaLeuAlaAlaAlaGlyAsnAspProPheLysArgThrThrAlaLeuGluThr 155

Qy     478 TGC------CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTA 531
                    ||||||:::||||||   |||         ::::::         |||
Db     156 PheSerLysGlnGluIleAlaLysCysGluGlu---GluGlnGlyTyrArgCysGlnPhe 174

Qy     532 GAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTT 591
                 |||         |||   |||   :::                     |||||||||
Db     175 PheSerPheAlaGlyGlyAsnThrTyrArgValPheLysAsnLeuGluIleLysAspVal 194

Qy     592 CGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGG 651
           |||:::|||:::||||||         |||||||||||||||||||||      |||||||||
Db     195 ArgLeuValTyrAlaProGlnGlySerValGlyLysPheGlyGlyAspValAspAsnTrp 214

Qy     652 ATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAAC 711
           |||||||||||||||||||||||||||      :::|||   |||   |||   |||
Db     215 MetTrpProArgHisThrGlyAspPheSerPheTyrArgAlaTyrValGlyLysAspGly 234

Qy     712 CGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTA 771
           :::||||||   :::||||||:::|||   ||||||:::|||   ::::::
Db     235 LysProAlaSerPheSerLysGluAsnIleProTyrArgProLysHisTrpLeuLysPhe 254

Qy     772 TCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACG 831
           |||   |||               |||:::   |||         |||:::||||||   |||
Db     255 SerAspGlnProLeuGlyAspGlyAspPheValMetValAlaGlyTyrProGlyArgThr 274

Qy     832 GATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGT 891
           :::||||||                                    |||   ||||||
Db     275 AsnArgTyrAlaLeuVal------------------AlaGluPheGluAsnThrAlaHis 288

Qy     892 ATCGAAGTTCGCGGTATCAAGCAA------------------GGCATCTGGAAGGAAGCC 933
                       |||   |||                     :::   :::   |||
Db     289 TrpThrTyrProValIleGlyGlnHisPheLysAsnLeuIleAlaLeuIleGluAlaAla 308

Qy     934 ATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAAC 993
                 :::         ::::::||||||||||||      |||         |||
Db     309 SerLysGlnAsnProAspIleGlnValLysTyrAlaSerThrLeuAlaGlyLeuAsnAsn 328

Qy     994 TATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGT 1053
                 ||||||      |||         |||         |||::::::   ||||||:::
Db     329 ThrSerLysAsnPheAspGlyGlnLeuAspGlyPheArgArgIleAsnAlaIleGlyGln 348

Qy    1054 AAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGC------------ 1101
           |||::::::||||||   |||         |||:::::::::   |||
Db     349 LysGlnSerGluGluThrAlaValLeuAlaTrpLeuLysGlnGlnGlyIleArgGlyHis 368

Qy    1102 AAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCC 1161
           ::::::            |||      |||   :::   |||
Db     369 GluAlaLeuAlaAlaHisGlnThrLeuValAspLeuThrGluGlnTyr------------ 384

Qy    1162 AAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTT 1221
           ||||||||||::::::         ::::::         |||   |||:::   |||:::
Db     385 LysAlaAsnGlnAspArgAspPheValLeuGlyGln---PheAsnGlySerGlyValIle 403

Qy    1222 CGTTTTGCACAGTTTGCCAACGCATTGGCT---------ACAAATCCTGATGCTCAT--- 1269
              |||         ||||||      |||         |||      ||||||
Db     404 GlyValAlaValAsnLeuTyrArgLeuAlaIleGluArgThrLysSerAspAlaGlnArg 423

Qy    1270 ------------------------------------GCCGGTATCCTCAAATCGCTTGAC 1293
                                                |||   ||||||   ::::::
Db     424 GluAlaGlyTyrGlnGluArgAspLeuProThrIleGluGlyAsnLeuLysGlnMetGlu 443

Qy    1294 GACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTG------------------ 1335
                 :::   |||||||||   :::||||||:::::
Db     444 ---------ArgArgTyrLeuProGluMetAspArgGlnMetGlnGlnTyrTrpLeuThr 460

Qy    1336 ------------CTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCC------ 1377
                       ||||||               |||::::::|||:::   |||
Db     461 GluTyrAsnLysLeuPro---------------ValLysGlnArgValAlaAlaIleAsp 475

Qy    1378 ------------GACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAA 1425
                       |||   :::|||      |||   :::   |||   |||
Db     476 ValTrpLeuGlyAspGlyIleProAlaThrLeuLysArgLeuGlyAspThrLysLeuSer 495

Qy    1426 GGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGAC 1485
              ::::::
Db     496 SerSerGluGluArg--------------------------------------------- 500

Qy    1486 AAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAA 1545
                          ||||||   :::   ::::::   |||               :::
Db     501 ---------------LeuLysTrpPheAsnAlaAspArgAlaAlaPheGluSerSerGln 515
```

```
Qy   1546 GATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGAT 1605
          |||||||||:::      :::   ::::::::   |||          |||:::
Db    516 AspProAlaIleArgTyrAlaValAlaIleMetProAlaLeuLeuGluIleGluArgGln 535

Qy   1606 GCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGT 1665
                               :::   |||   :::   ::::::         |||
Db    536 AsnLysIleArgThrGlyGluLeuLeuLysAlaArgProIleTyrLeuGlnAlaLeuAla 555

Qy   1666 GAGATGTACCCC------GGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATG 1719
          :::                |||:::   :::     |||||||||   ::::::|||:::
Db    556 AspTyrAsnLysSerHisGlyLysPheValTyrProAspAlaAsnSerSerLeuArgIle 575

Qy   1720 AGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACG 1779
          ::::::|||   ::::|||||||||   |||:::||||||         |||       |||
Db    576 ThrPheGlyHisValLysGlyTyrSerProLysAspGlyValGluTyrThrProPheThr 595

Qy   1780 ACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAG 1839
          |||   :::||||||:::   |||      :::          :::   |||         :::
Db    596 ThrLeuGlnGlyValMetAlaLys---AsnThrGlyValGluProPheAspSerProLys 614

Qy   1840 AATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAAC------GGT 1893
          ::::::::::::      :::   |||:::|||          |||:::            |||
Db    615 SerLeuIleAsnAlaIleLysAlaLysSerTyrAlaAsnLeuAlaAspGlnArgIleGly 634

Qy   1894 CAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCC 1953
             :::   :::   |||||||||:::   |||||||||||||||||||||||||||
Db    635 ThrValProValAsnPheLeuSerAspLeuAspIleThrGlyGlyAsnSerGlySerPro 654

Qy   1954 GTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATG 2013
          |||   |||   :::|||:::|||||||||||||||||||||||||||||||||:::::
Db    655 ValLeuAspAlaHisGlyLysLeuValGlyLeuAlaPheAspGlyAsnTrpGluSerVal 674

Qy   2014 AGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTAC 2073
          |||   :::      |||:::|||   :::   |||||||||:::||||||   ||||||
Db    675 SerSerAsnTrpValPheAspProValMetThrArgThrIleAlaValAspSerArgTyr 694

Qy   2074 GTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTG 2133
          |||   :::::::::   :::               |||   |||::::::||||||   |||
Db    695 ValGlnTrpIleMetThrGluValAlaProAlaProHisLeuLeuLysGluLeuAsnLeu 714
```

```
RESULT 2
Q8WWQ4
ID   Q8WWQ4      PRELIMINARY;      PRT;  1349 AA.
AC   Q8WWQ4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Mucin 5 (Fragment).
GN   MUC5AC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21426417; PubMed=11535137;
RA   Escande F., Aubert J.P., Porchet N., Buisine M.P.;
RT   "Human mucin gene MUC5AC: organization of its 5'-region and central
RT   repetitive region.";
RL   Biochem. J. 358:763-772(2001).
DR   EMBL; AJ298318; CAC83675.1; -.
FT   NON_TER       1      1
FT   NON_TER    1349   1349
SQ   SEQUENCE   1349 AA;  135600 MW;  4DC3C1544F1E5EBA CRC64;

Alignment Scores:
Pred. No.:              1.21e-05      Length:         1349
Score:                  206.00        Matches:        168
Percent Similarity:     32.71%        Conservative:   78
Best Local Similarity:  22.34%        Mismatches:     281
Query Match:            5.39%         Indels:         225
DB:                     4             Gaps:           30
```

```
US-10-008-355-1 (1-2139) x Q8WWQ4 (1-1349)

Qy     41 TGCTGTTGGGTGCTTCAGGGGTAGCCA--------------------AAGCCGACAAAG 79
          ||||||         :::|||   |||                      |||:::
Db    593 CysCysGluThrProLysGlyCysProValThrSerThrSerValThrAlaProSerPro 612

Qy     80 GCATGTGGC-----------------TCCTCAACGAACTCAATCAGGAGAATCTGGATC 121
             |||                     ||||||:::|||                |||
Db    613 LeuValGlyGluProProAlaGlnThrGlnSerThrSerSer-----------TrpGln 628

Qy    122 GAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGC--- 178
          :::      ::::::   |||   :::         |||   ||||||      ||||||
Db    629 LysSerArgThrThrThrLeuValThrSerSerIleThrSerThrThrGlnThrSerThr 648

Qy    179 -----------------------CGTCCATTGCCAATGCCGTGGTTATCTTCGGTGGCG 214
                                    |||      :::|||      :::|||
Db    649 ThrSerAlaProThrThrSerThrThrProAlaSerIleProSerThrThrSerAlaPro 668

Qy    215 GATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCACTGCGGAT 274
                         :::      |||          :::         ||||||:::|||
Db    669 ThrThrSerThrThrSerAlaPro-----------ThrThrSerThrThrSerAlaPro 684

Qy    275 ACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTT 334
          |||      |||                  ||||||      |||      |||
Db    685 ThrThr---SerThrThrSerThrProGlnThrThrThrSerSerAlaProThrSerSer 703

Qy    335 CTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGA 394

Db    703 ------------------------------------------------------------ 703

Qy    395 TCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGC 454
                         |||            :::|||   :::|||               |||
Db    704 -----------ThrThrSerAlaProThrThrSerThrIleSerAlaProThrThrSer 719

Qy    455 GTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGA 514
                :::                     :::   |||                 :::
Db    720 ThrIleSerAlaProThrThrSerThrThrSerAlaProThrAlaSerThrThrSerAla 739

Qy    515 ACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACG 574
                |||      |||            |||||||||||||||:::|||:::   |||
Db    740 ProThrSerThrSer---------SerAlaProThrThrAsnThrThrSerAlaProThr 756

Qy    575 ATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCA---GCTCTGTAGGTAAGTTCG 631
                :::   :::                           |||         :::|||:::
Db    757 ThrSerThrThrSer-----------------AlaProIleThrSerThrIleSerAla 770

Qy    632 GAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCG 691
                      ||||||      :::   :::      :::|||      ::::::
Db    771 ProThrThrSerThrThrSerThrProGlnThrSerThrIleSerSerProThrThrSer 790

Qy    692 TGTATGCCGGTGCCG---------ACAACCGGCCGGCCGAATACAGCAAGGACAATAAAC 742
                |||   |||               ||||||      |||   |||:::   |||
Db    791 ThrThrProThrProGlnThrSerThrThrSerSerProThrThrSerThrThrSerAla 810

Qy    743 CCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATG 802
          |||   :::         ||||||   |||                  |||      |||:::
Db    811 ProThrThrSerThrThrSerAlaPro-----------ThrThrSerThrThrSerThr 826

Qy    803 CCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAG 862
          |||      |||:::|||   |||            :::|||:::
Db    827 ProGlnThrSerIleSerSerAlaProThrSerSerThrThrSerAlaProThrAlaSer 846

Qy    863 ATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCT 922
             :::|||         |||         |||            |||:::
Db    847 ThrIleSerAlaProThrThrSerThrThrSerPheHisThrThrSerThr--------- 863

Qy    923 GGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTC 982

Db    863 ------------------------------------------------------------ 863

Qy    983 AGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACG 1042
```

```
Db    863 ------------------------------------------------------------ 863

Qy   1043 TGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCA 1102
                    |||  |||             |||  |||  |||
Db    864 ---------ThrSerProProThr---------SerSerThrSerSerThrProGlnThr 877

Qy   1103 AGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCA 1162
                   |||  |||               |||                      |||
Db    878 SerLysThrSerAlaAlaThrSerSerThrThrSerGlySerGlyThrThrProSerPro 897

Qy   1163 AGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTC 1222
            ||||||
Db    898 ValProThrThr------------------------------------------------ 901

Qy   1223 GTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCA 1282
                                                   :::       |||:::
Db    902 -----------------------------SerThrAlaSerValSerLysThrSerThr 911

Qy   1283 AATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCG 1342
          :::   ::::::  |||      |||  |||          |||      |||  |||
Db    912 SerHisValSerValSerLysThrThrHisSerGlnProValThrArgAspCysHisPro 931

Qy   1343 CCATGCTCG---ATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCG--------- 1390
             |||:::          :::       |||     |||:::      |||
Db    932 ArgCysThrTrpThrLysTrpPheAspValAspPheProSerProGlyProHisGlyGly 951

Qy   1391 ---------ATATATTCAAGAATGTAATCGA------CAAGAAATTCAAAGGCGACACGA 1435
                   |||:::    :::      |||      ::::::|||   ||||||    ||
Db    952 AspLysGluThrTyrAsnAsnIle-IleArgSerGlyGluLysIleCysArgArgProGl 971

Qy   1436 AGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATG 1495
          ||||:::   |||||||:::   |||   |||
Db    971 uGluIleThrArgLeuGlnCysArgAlaGlu--------------------------Se 982

Qy   1496 CCATGCTCAAGTCCATGGACAA------GGAAAAGTTTGCCAAGGCTATCGAGAAAGATC 1549
             |||   :::|||        :::       |||:::|||   |||
Db    982 rHisProGluValSerIleGluHisLeuGlyGlnValValGln----------------- 996

Qy   1550 CGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGA 1609
                                           |||||||||      |||
Db    997 ---------------------------------CysSerArgGluGluGly-------- 1002

Qy   1610 TGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGA 1669
                    :::|||        :::   |||                    :::|||:::
Db   1003 -------LeuValCysArgAsnGlnAspGlnGlnGly--ProPheLysMetCysLeuAsn 1019

Qy   1670 TGTACCCCGGACGTGCTCTGCCGAGCGATGCCA--------------ACTTCACCATGC 1714
                   |||||||||          |||                |||||||
Db   1020 TyrGluValArgValLeuCysCysGluThrProLysGlyCysProValThrSerThrPro 1039

Qy   1715 GTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATC 1774
          |||       ||||||||||||                |||   |||   |||
Db   1040 Val------ThrAlaProSerThrProSerGlyArgAlaThrSerPro------------ 1053

Qy   1775 ATACGACAGGCAAGGGCGTATTGG---AGAAGCAGGATCCTAAGAGCGATGAGTTTGCCG 1831
                |||:::    :::    |||    :::||||||               ::::::
Db   1054 ---ThrGlnSerThrSerSerTrpGlnLysSerArgThrThrThrLeuValThr------ 1070

Qy   1832 TACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACG 1891
                     :::||||||:::|||    |||:::|||            |||
Db   1071 -----------ThrSerThrThrSerThrProGlnThrSerThrThrSerAlaProThr 1086

Qy   1892 GTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTA--- 1948
             |||:::|||       |||      :::|||||||||          |||
Db   1087 ThrSerThrIleProAlaSerThrProSerThrThrSerAlaProThrThrSerThrThr 1106

Qy   1949 ---GCCCCGTATTCGATAAGAACGGCC------------------------------GTC 1975
             ||||||    :::       |||:::
Db   1107 SerAlaProThrThrSerThrThrSerAlaProThrHisArgThrThrSerGlyProThr 1126

Qy   1976 TGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAAC 2035
             |||      |||:::        |||                  |||:::|||:::|||
Db   1127 ThrSerThrThrLeuAlaProThrThrSerThrThrSerAlaProThrThrSerThrAsn 1146

Qy   2036 ------CCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTG 2089
                |||      |||       ||||||    |||:::::|||    |||:::
Db   1147 SerAlaProThrThrSerThrIleSerAlaSerThrThrSerThrIleSerAlaProThr 1166

Qy   2090 ACAAATGGGGTCAGTGCCCCCGTCTCATCC 2119
          |||:::    :::|||:::|||    ||||||
Db   1167 ThrSerThrIleSerSerProThrSerSer 1176

RESULT 3
Q8TFG4
ID   Q8TFG4      PRELIMINARY;      PRT;   800 AA.
AC   Q8TFG4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative glucoamylase, similar to S. cerevisiae STA1, similar to S.
DE   pombe SPBPJ4664.02.
GN   SPAPB18E9.04C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   O'Neil S., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL691402; CAD27470.1; -.
SQ   SEQUENCE   800 AA;  79351 MW;  F65488E43FE7089A CRC64;

Alignment Scores:
Pred. No.:               0.000984      Length:        800
Score:                   177.00        Matches:       185
Percent Similarity:      34.78%        Conservative:  111
Best Local Similarity:   21.74%        Mismatches:    315
Query Match:             4.63%         Indels:        241
DB:                      3             Gaps:          35

US-10-008-355-1 (1-2139) x Q8TFG4 (1-800)

Qy     34 GCAGCCCTGCTGTTGGGTGCTTCAGGGGTAGCCAAAGCCGACAAAGGCATGTGGC----- 88
          :::   |||||||||   |||          :::|||   |||       :::||||||
Db      7 SerThrLeuLeuLeuLeuAlaGlnHisLeuAlaProAla-AlaAlaSerCysGlyLeuAl 26

Qy     89 ------------TCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTC 135
                      |||   |||   ::::::   |||:::   :::    |||:::::
Db     26 aAsnGluAlaSerLeuSerAlaAsnGluValGlnAspIleTyrHisAspPheValAlaTh 46

Qy    136 GGCTTTA--------------------CGCTCCCGTTGGATTCGCTCTACAGTTTCGAC 174
          :::::::                    |||:::       :::    :::::::::|||||
Db     46 rSerIleAsnTyrAlaAspLeuLeuLysArgAsnGluAspLeuAsnAlaSerLeuSerTh 66

Qy    175 AAGCCGTCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTG 234
          |     |||:::    :::                |||          |||          |||
Db     66 rGlySerProValGluIleThrSerThrSerCysThrThrAspThrSerAlaSer---Th 85

Qy    235 TCCGATCAGGGCCTGATCT----------------------------------------- 253
          ||||||    |||
Db     85 rProIleIleThrGluSerThrSerSerThrSerSerAlaSerThrThrGlySerSerSe 105

Qy    254 -------TTACCAACCACC------------ACTGCGGATACGGTGCTATCCAGAGCCAA 294
                 ||||||:::|||            |||       |||    :::  |||   |||
Db    105 rSerProLeuProSerThrSerThrSerCysThrThrSerThrSerIle-Pro---ProT 124

Qy    295 AGCACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAG 354
                |||    |||    ||||||                                   :
Db    124 hrGlyGlySerSerSerLeuSerThrProIleThrProThrValProProThrSerThrS 144

Qy    355 CTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAG 414
          ::|||    |||                   |||             :::   |||
Db    144 erSerThrSerIleProIleProProThrSerThrSerSerThrAspThrAsnSerAsnP 164
```

```
Qy     415 GTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAG 474
                   |||                         |||     :::|||
Db     164 roLeuProThrThr--------------------SerThrSerCysThrThrSerThrS 177

Qy     475 GTATGCCAAGAACTGGCCA-AAAAAGAAAATGCAGACG--------------------A 512
            :::|||   ||||||              :::   |||
Db     177 erIleProProThrGlyGlySerSerSerLeuSerThrProIleThrProThrValProP 197

Qy     513 GAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTA 572
            |||::::::::|||   |||:::   ||||||   |||:::||||||||:::   |
Db     197 roThrSerThrSerSerThrSerIleProIleProProThrSerThrSerSerThrAspT 217

Qy     573 CGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGT---- 628
           ||       |||                     ||||||             |||
Db     217 hrAsnSerSerPro--------------------LeuProThrThrSerThrSerCysT 230

Qy     629 -----TCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGG------------ 670
                 |||   :::|||   |||   |||          ::::::
Db     230 hrThrSerThrSerIleProThrGlyGlySerSerSerLeuSerThrProIleThrProT 250

Qy     671 --------GCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCG--------------- 706
                   ||||||      |||       :::|||:::|||
Db     250 hrValProProThrSerThrSerSerThrSerIleProIleProProThrSerThrSerS 270

Qy     707 --------------------------------------------------ACAACCG 713
                                                             :::|||
Db     270 erThrAspThrAsnSerSerProLeuProThrThrSerThrSerCysThrThrSerThrS 290

Qy     714 GCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCG------------------ 754
             |||   |||       :::      |||:::::|||
Db     290 erIleProProThrGlyAsnSerThrThrProValThrProThrValProProThrSerT 310

Qy     755 -----TTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACG---------ACTA 800
                 ||||||   |||   |||      :::|||      |||          :::
Db     310 hrSerSerThrSerThrProProProProAlaSerThrSerSerThrGlyThrSerSerS 330

Qy     801 TGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGA 860
             |||   ||||||   |||   :::         :::||||||:::
Db     330 erProLeuProSerThrSerThrSerCysThrThrSerThrSerIleProProThrGlyA 350

Qy     861 AGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCAT 920
                 :::         |||   |||      :::   ||||||   ::::
Db     350 snSerThrThrProVal---ThrProThrValProProThrSerThrSerSerThrSerT 369

Qy     921 CTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGC 980
           ::         |||   :::                  |||:::   |||
Db     369 hrProProProAlaSerThrSerSerThrGlyThrSerSerSerProLeu------- 386

Qy     981 TCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGA 1040
                                            :::   |||:::   |||         :
Db     387 ------------------------------LeuSerThrSerThrSerCysThrThrS 396

Qy    1041 CGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGG 1100
           ::               |||:::|||         :::|||   |||         |||
Db     396 er---------ThrSerIleProProThrGlyAsnSerThrThrProValThrProThrV 413

Qy    1101 CAAGAGTGCTGTCTATGGCG---------------ATGTATTGTCTTCTCTCGAAAAGGC 1145
                       |||   :::                 |||      |||        :
Db     413 alProProThrSerSerSerThrProLeuThrThrThrAsnCysThrThrSerThrSerV 433

Qy    1146 TT-----------------ATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTT 1187
           ::                  :::         |||   |||
Db     433 alProTyrThrSerThrProValThrSerThrProLeuAlaThrThrAsn---------- 449

Qy    1188 GAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATT 1247
                       |||:::   ||||||
Db     450 --CysThrThrSerThrSerValPro-------------------------------- 457

Qy    1248 GGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGA 1307
                                 ||||||:::|||      ||||||||:::          |
Db     458 ----------TyrThrSerThrProValThrSerThrProLeuThrThrThrAsnCysT 474

Qy    1308 CTACCTCC--------------------------------------CCTCGCTCGACCG 1328
           ||||||||                                      |||   :::|||
Db     474 hrThrSerThrSerIleProTyrThrSerThrProValThrSerThrProLeuThrThrT 494

Qy    1329 TAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCC 1388
               |||            |||:::   |||         |||   |||   :::|||
Db     494 hrAsnCysThrThrSerThrSerValProTyr------ThrSerThrProValThrSerS 512

Qy    1389 CGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGA 1448
                  |||:::         ||||||               |||         :::   |
Db     512 erAsnTyrThrIleSerSerSerThr---------ProValThrSerThrProValThrT 529

Qy    1449 CTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTC 1508
            ||:::      |||               ::::::   ||||||:::         |||:::|
Db     529 hrThrAsnCysThrThrSerThrSerValLeuTyrThrSerThrProValThrSerThrP 549

Qy    1509 CA-----------------TGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCC 1550
           ||                   |||      |||:::|||      |||      :::
Db     549 roLeuAlaThrThrAsnCysThrThrSerThrSerValProTyrThrSerThrProValT 569

Qy    1551 GGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGAT 1610
                  ::::::   :::
Db     569 hrSerSerAsnTyrThrIleSer------------------------------------- 576

Qy    1611 GGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTT---TGCGTGA 1667
                                         |||   |||   ||||||
Db     577 -------------------SerSerThrProValThrSerThrProValThrThrThrA 590

Qy    1668 GATGTACC------CCGGACGTGCTCTGCCGAGCGATGCCA---ACTTCACCA------- 1711
              ||||||         ||||||      :::   |||   |||||||||
Db     590 snCysThrThrSerThrSerValLeuTyrThrSerThrProIleThrSerProAsnSerT 610

Qy    1712 -----------------------------------------TGCGTATGAGCTACGGCTC 1730
                                                       :::      |||:::
Db     610 hrSerSerSerSerThrGlnValSerTrpAsnSerThrThrProIleThrGlyThrSerT 630

Qy    1731 CATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGG 1790
            |||:::      :::      ::::::|||   |||:::         |||   :::
Db     630 hrSerLysValThrSerSerThrSerIleProLeuThrSerThrAsnArgThrSerThrT 650

Qy    1791 CGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCG------------------ 1831
             :::      |||      :::         |||
Db     650 hrPheThrSerSerThrSerIleSerThrSerSerSerSerThrAlaThrSerSerThrS 670

Qy    1832 -------------------------TACAGGAGAATA---------TCCTCGACCTCTT 1856
                                    |||      :::         |||||||||   |
Db     670 erPheAlaSerGluSerSerSerPheTyrSerAsnValThrThrSerSerSerThrValS 690

Qy    1857 CCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATC 1916
            ||   |||               :::   |||            |||:::   :::   |||:::
Db     690 erThrProProProThrThrSerPhePro---------SerThrPheThrThrSerPheI 707

Qy    1917 GAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCT 1976
            |||::::::|||   :::         |||               ::::::||||||
Db     707 leThrSerSerSerLeuSerSerIleProAsnAsnSerThrGluValLysThrAla---- 725

Qy    1977 GATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACC 2036
                         |||   ::::::|||         :::   |||:::|||::::::
Db     726 --------------SerThrSerSerGlyThrGlu---IleLysThrAlaSerThrSerS 740

Qy    2037 CGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATG 2096
                   |||:::   |||::::::|||   |||:::   :::|||      |||:::
Db     740 erGlySerSerSerSerSerSerTyrThrProAlaSerSerThrSerThrThrThrSerS 760

Qy    2097 GGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
              ||||||:::      ||||||   |||
Db     760 erValSerSerArgGlnSerSerSerSer 769
```

RESULT 4
O14651
ID O14651 PRELIMINARY; PRT; 957 AA.
AC O14651; O14650;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)

```
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Mucin (Intestinal mucin) (Fragment).
GN   MUC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97476275; PubMed=9334251;
RA   Gum J.R. Jr., Ho J.J.L., Pratt W.S., Hicks J.W., Hill A.S.,
RA   Vinall L.E., Roberton A.M., Swallow D.M., Kim Y.S.;
RT   "MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl
RT   terminus, and a novel upstream repetitive region.";
RL   J. Biol. Chem. 272:26678-26686(1997).
DR   EMBL; AF007194; AAC02272.1; -.
DR   EMBL; AF007196; AAB84383.1; -.
DR   EMBL; AF007195; AAB84382.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; 1.
KW   EGF-like domain; Glycoprotein.
FT   NON_TER       1      1
SQ   SEQUENCE   957 AA;  99554 MW;  E67722560D600A1F CRC64;

Alignment Scores:
Pred. No.:              0.0267        Length:         957
Score:                  157.00        Matches:        170
Percent Similarity:     33.00%        Conservative:   93
Best Local Similarity:  21.33%        Mismatches:     289
Query Match:            4.11%         Indels:         245
DB:                     4             Gaps:           32

US-10-008-355-1 (1-2139) x O14651 (1-957)

Qy     155 ATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTTCGGTGGCG 214
               |||   :::|||   :::||||||          |||          |||      :::
Db     154 IleThrThrThrGluThrThrSerHisAspThrPro------------SerPheThrSer 169

Qy     215 GATGTACCGGTATCACAGTGTCCGATCAGG------------GCCTGATCTTTACCAACC 262
              :::      |||:::   |||                :::   ||||||    |||
Db     170 SerIleThrThrSerGluThrProSerHisSerThrProSerSerThrSerLeuIleThr 189

Qy     263 ACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCG 322
           ||||||      |||      |||                 |||:::
Db     190 ThrThrLysThrThrSerHisSerThrProSer---------PheThrSerSerIleThr 206

Qy     323 ATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGT 382
                         :::      |||||||||
Db     207 ThrThrGluThrThrSerHisSerAlaArgSerPhe------------------------ 218

Qy     383 ATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTG 442
                                                    :::|||   ::::::
Db     219 ------------------------------------------ThrSerSerIleThrThr 224

Qy     443 ACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAA 502
           |||      :::                            :::::::::
Db     225 ThrGluThrThr--------------SerHisAsnThrArgSerPheThrSerSerIle 239

Qy     503 ATGCAGACGAGAACCAACTCT---------GCATCGTAGAGCCTTTCTATTCCAACAACG 553
                 |||  |||||||||            |||   :::   ||||||   ||||||
Db     240 ThrThrThrGluThrAsnSerHisSerThrThrSerPheThrSerSerIleThrThrThr 259

Qy     554 AATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCA 613
            |||:::|||   ||||||                                    |||
Db     260 GluThrThrSerHisSerThr-----------------------------------Pro 267

Qy     614 GCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACT------------------------ 649
           :::           |||   :::|||   ||||||
Db     268 SerPhe---------SerSerSerIleThrThrThrGluThrProLeuHisSerThrPro 284

Qy     650 GGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACA 709
           |||                 |||          |||          |||       |||
Db     285 GlyLeuProSerTrpValThrThrThrLysThrThrSerHisIleThrProGlyLeuThr 304

Qy     710 ACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCG 769
           :::           |||       |||   :::          |||||||||
Db     305 SerSerIleThrThrThrGluThrThrSerHisSerThrProGlyPheThrSer------ 322

Qy     770 TATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTA 829
                        |||       |||:::         ||||||:::|||   :::
Db     323 ---SerIleThrThrThrGluThrThrSerGluSerThrProSerLeuSerSerSerThr 341

Qy     830 CGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTC 889
                                  ::::::      ||||||   :::   |||   :::
Db     342 ----------------------IleTyrSerThrValSerThrSerThrThrAlaIle 353

Qy     890 GTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGG 949
              |||   |||       |||      |||             |||
Db     354 ThrSerHisPheThrThrSerGluThrAlaValThrProThrProValThrProSerSer 373

Qy     950 CTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGA 1009
           |||      :::::::::|||   :::   ||||||   ||||||
Db     374 LeuSer---ThrAspIleProThrThrSerLeuArgThrLeuThrProSer--------- 389

Qy    1010 TCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTG--------- 1060
           ||||||   |||:::   ||||||       |||            |||:::
Db     390 SerValGlyThrSerThrSerLeuThrThrThrThrAspPheProSerIleProThrAsp 409

Qy    1061 -----------CCGAGGAAAGAGCAT------TCGCAGACTGGATCCGTAAGAACGGCA 1102
                      |||   :::   |||      |||   :::   |||:::::::::
Db     410 IleSerThrLeuProThrArgThrHisIleIleSerSerSerProSerIleGlnSerThr 429

Qy    1103 AGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCA 1162
                    |||:::                  :::          :::|||
Db     430 GluThrSerSerLeuValGlyThrThrSerProThrMetSerThrValArgMetThrLeu 449

Qy    1163 AGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTC 1222
           |||   |||          |||   :::     :::|||:::|||
Db     450 ArgIleThrGluAsnThrProIleSerSerPheSerThrSerIleVal------------ 465

Qy    1223 GTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCA 1282
           |||:::         ||||||                           ||||||   :::
Db     466 ValIleProGluThrProThr-----------------GlnThrProProValLeuThr 479

Qy    1283 AATCGCTTGACGACAAGTACA------------------------------AAGACTACC 1312
           :::         |||   |||                                    |||
Db     480 SerAlaThrGlyThrGlnThrSerProAlaProThrThrValThrPheGlySerThrAsp 499

Qy    1313 TCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCC 1372
           |||   ||||||:::                |||   :::   |||         ||||||
Db     500 SerSerThrSerThrLeuHisThrLeuThrProSerThrAlaLeuSerThrIleValSer 519

Qy    1373 ---------------CTGCCG------------------------------------- 1378
                         :::|||
Db     520 ThrSerGlnValProIleProSerThrHisSerSerThrLeuGlnThrThrProSerThr 539

Qy    1379 --------------------ACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGA 1417
                               ||||||:::         ::::::          ||||||
Db     540 ProSerLeuGlnThrSerLeuThrSerThrSerGluPheThrThrGluSerPheThrArg 559

Qy    1418 AATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTC--- 1474
              |||   :::|||      ::::::   ||||||:::||||||   :::|||
Db     560 GlySerThrSerThr---AsnAlaIleLeuThrSerPheSerThrIleIleTrpSerSer 578

Qy    1475 ------CTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTG 1528
                 |||         ||||||   |||   |||:::          |||      |||:
Db     579 ThrProThrIleIleMetSerSerSerProSerSerAlaSerIleThr--ProValPheS 598

Qy    1529 CCAAGGCTATCGAGAAAGATCCGGCA--------------------------------- 1554
           ::        |||       |||:::
Db     598 erThrThrIleHisSerValProSerSerProTyrIlePheSerThrGluAsnValGlyS 618

Qy    1555 -------------------GTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTA 1594
```

```
                                 ||||||   |||          :::      :::
Db    618 erAlaSerIleThrGlyPheProSerLeuSerSerSerAlaThrThrSerThrSerSerT 638

Qy   1595 TTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTG 1654
             :::                                |||         |||      ||||||
Db    638 hrSerSerSerLeuThrThrAlaLeuThrGluIleThrPro-PheSerTyrIleSerLeu 657

Qy   1655 CCGGTTTGCGTGAGATGTACCCCGGACGTG---------CTCTGCCGAGCGATGCCAACT 1705
          |||          |||          :::       :::      |||    ||||||
Db    658 ProSerThrThrProCysProGlyThrIleThrIleThrIleValProAlaSerProThr 677

Qy   1706 TCACCATGCGTA------------------TGAGCTACGGCTCCA--------------- 1732
             |||||||||                      ||||||:::|||
Db    678 AspProCysValGluMetAspProSerThrGluAlaThrSerProProThrThrProLeu 697

Qy   1732 ------------------------------------------------------------ 1732

Db    698 ThrValPheProPheThrThrGluMetValThrCysProThrSerIleSerIleGlnThr 717

Qy   1732 ------------------------------------------------------------ 1732

Db    718 ThrLeuThrThrTyrMetAspThrSerSerMetMetProGluSerGluSerSerIleSer 737

Qy   1733 ---------TCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAG 1783
                   |||                |||||||||      |||:::
Db    738 ProAsnAlaSerSerSerThrGlyThrGlyThrValProThrAsnThrValPheThrSer 757

Qy   1784 GCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATA 1843
             |||                                   ||||||
Db    758 ThrArg------------------------------------LeuPro---------Thr 762

Qy   1844 TCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACG---GTCAGCTCC 1900
          |||   |||         :::       :::::::::    :::|||        ||||||:::
Db    763 SerGluThrTrpLeuSerAsnSerSerValIleProLeuProLeuProGlyValSerThr 782

Qy   1901 ATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCG 1960
          |||   |||:::   :::    ::::::|||                    |||
Db    783 IleProLeuThrMetLysProSerSerSer------------------LeuProThrIle 796

Qy   1961 ATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTG 2020
          :::||||||:::          |||          |||     |||    :::
Db    797 LeuArgThrSerSerLysSerThrHisProSerProProThrThrArg------------ 812

Qy   2021 ACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCT 2080
              :::|||      |||:::    :::    |||          |||        |||    :::
Db    813 ---ThrSerGluThrProValAlaThrThrGln---------ThrProThrThrLeuThr 828

Qy   2081 TCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
          |||       |||       :::::::::   :::::::::::|||
Db    829 SerArgArgThrThrArgIleThrSerGlnMetThrThrGlnSer 843
```

RESULT 5
Q9UKW9

```
ID   Q9UKW9      PRELIMINARY;      PRT;  1217 AA.
AC   Q9UKW9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Intestinal mucin 3 (Fragment).
GN   MUC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RX   MEDLINE=99443732; PubMed=10512748;
RA   Crawley S.C., Gum J.R. Jr., Hicks J.W., Pratt W.S., Aubert J.P.,
RA   Swallow D.M., Kim Y.S.;
RT   "Genomic organization and structure of the 3' region of human MUC3:
RT   alternative splicing predicts membrane-bound and soluble forms of the
RT   mucin.";
RL   Biochem. Biophys. Res. Commun. 263:728-736(1999).
DR   EMBL; AF113616; AAF13032.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF01390; SEA; 1.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   EGF-like domain; Glycoprotein.
FT   NON_TER      1      1
SQ   SEQUENCE   1217 AA;  129171 MW;  0783A4003C969664 CRC64;
```

Alignment Scores:

Pred. No.:	0.0302	Length:	1217
Score:	157.00	Matches:	170
Percent Similarity:	33.00%	Conservative:	93
Best Local Similarity:	21.33%	Mismatches:	289
Query Match:	4.11%	Indels:	245
DB:	4	Gaps:	32

US-10-008-355-1 (1-2139) x Q9UKW9 (1-1217)

```
Qy    155 ATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTTCGGTGGCG 214
          |||    :::|||    :::||||||          |||          |||       :::
Db    154 IleThrThrThrGluThrThrSerHisAspThrPro-----------SerPheThrSer 169

Qy    215 GATGTACCGGTATCACAGTGTCCGATCAGG------------GCCTGATCTTTACCAACC 262
             :::         |||:::   |||                :::   ||||||   |||
Db    170 SerIleThrThrSerGluThrProSerHisSerThrProSerSerThrSerLeuIleThr 189

Qy    263 ACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCG 322
          ||||||      |||        |||                      |||:::
Db    190 ThrThrLysThrThrSerHisSerThrProSer---------PheThrSerSerIleThr 206

Qy    323 ATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGT 382
                              :::        |||||||||
Db    207 ThrThrGluThrThrSerHisSerAlaArgSerPhe----------------------- 218

Qy    383 ATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTG 442
                                                    :::|||    ::::::
Db    219 -----------------------------------------ThrSerSerIleThrThr 224

Qy    443 ACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAA 502
          |||       :::                                  :::::::::
Db    225 ThrGluThrThr---------------SerHisAsnThrArgSerPheThrSerSerIle 239

Qy    503 ATGCAGACGAGAACCAACTCT---------GCATCGTAGAGCCTTTCTATTCCAACAACG 553
                |||   |||||||||             |||    :::   ||||||   ||||||
Db    240 ThrThrThrGluThrAsnSerHisSerThrThrSerPheThrSerSerIleThrThrThr 259

Qy    554 AATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCA 613
             |||:::|||   ||||||                                   |||
Db    260 GluThrThrSerHisSerThr-----------------------------------Pro 267

Qy    614 GCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACT------------------------ 649
          :::          |||   :::|||   ||||||
Db    268 SerPhe---------SerSerSerIleThrThrThrGluThrProLeuHisSerThrPro 284

Qy    650 GGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACA 709
          |||                  |||          |||          |||       |||
Db    285 GlyLeuThrSerTrpValThrThrThrLysThrThrSerHisIleThrProGlyLeuThr 304

Qy    710 ACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCG 769
          :::            |||      |||   :::          |||||||||
Db    305 SerSerIleThrThrThrGluThrThrSerHisSerThrProGlyPheThrSer------ 322

Qy    770 TATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTA 829
                      |||      |||:::      ||||||:::|||      :::
Db    323 ---SerIleThrThrThrGluThrThrSerGluSerThrProSerLeuSerSerSerThr 341

Qy    830 CGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTC 889
```

```
                                       ::::::    ||||||   :::   |||   :::
Db    342 -----------------------IleTyrSerThrValSerThrSerThrThrAlaIle 353

Qy    890 GTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGG 949
              |||   |||      |||     |||        |||
Db    354 ThrSerHisPheThrThrSerGluThrAlaValThrProThrProValThrProSerSer 373

Qy    950 CTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGA 1009
          |||      :::::::|||   :::   ||||||   ||||||
Db    374 LeuSer---ThrAspIleProThrThrSerLeuArgThrLeuThrProSer--------- 389

Qy   1010 TCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTG--------- 1060
          ||||||   |||:::   ||||||    |||           |||:::
Db    390 SerValGlyThrSerThrSerLeuThrThrThrThrAspPheProSerIleProThrAsp 409

Qy   1061 -----------CCGAGGAAAGAGCAT------TCGCAGACTGGATCCGTAAGAACGGCA 1102
                        |||   :::   |||     |||   :::   |||:::::::::
Db    410 IleSerThrLeuProThrArgThrHisIleIleSerSerSerProSerIleGlnSerThr 429

Qy   1103 AGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCA 1162
                   |||:::                :::      :::|||
Db    430 GluThrSerSerLeuValGlyThrThrSerProThrMetSerThrValArgMetThrLeu 449

Qy   1163 AGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTC 1222
          |||   |||          |||   :::      :::|||:::|||
Db    450 ArgIleThrGluAsnThrProIleSerSerPheSerThrSerIleVal------------ 465

Qy   1223 GTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCA 1282
          |||:::         ||||||                        ||||||   :::
Db    466 ValIleProGluThrProThr------------------GlnThrProProValLeuThr 479

Qy   1283 AATCGCTTGACGACAAGTACA------------------------------AAGACTACC 1312
          :::          |||   |||                            .|||
Db    480 SerAlaThrGlyThrGlnThrSerProAlaProThrThrValThrPheGlySerThrAsp 499

Qy   1313 TCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCC 1372
          |||      ||||||:::                |||   :::   |||        ||||||
Db    500 SerSerThrSerThrLeuHisThrLeuThrProSerThrAlaLeuSerThrIleValSer 519

Qy   1373 ---------------CTGCCG------------------------------------- 1378
                         :::|||
Db    520 ThrSerGlnValProIleProSerThrHisSerSerThrLeuGlnThrThrProSerThr 539

Qy   1379 --------------------ACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGA 1417
                              ||||||:::        ::::::           ||||||
Db    540 ProSerLeuGlnThrSerLeuThrSerThrSerGluPheThrThrGluSerPheThrArg 559

Qy   1418 AATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTC--- 1474
             |||   :::|||      ::::::   ||||||:::||||||   :::|||
Db    560 GlySerThrSerThr---AsnAlaIleLeuThrSerPheSerThrIleIleTrpSerSer 578

Qy   1475 ------CTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTG 1528
                   |||      ||||||   |||   |||:::      |||       |||:
Db    579 ThrProThrIleIleMetSerSerSerProSerSerAlaSerIleThr--ProValPheS 598

Qy   1529 CCAAGGCTATCGAGAAAGATCCGGCA-------------------------------- 1554
          ::       |||          |||:::
Db    598 erThrThrIleHisSerValProSerSerProTyrIlePheSerThrGluAsnValGlyS 618

Qy   1555 -------------------GTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTA 1594
                                   ||||||   |||          :::      :::
Db    618 erAlaSerIleThrGlyPheProSerLeuSerSerSerAlaThrThrSerThrSerSerT 638

Qy   1595 TTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTG 1654
               :::                        |||        |||      ||||||
Db    638 hrSerSerSerLeuThrThrAlaLeuThrGluIleThrPro-PheSerTyrIleSerLeu 657

Qy   1655 CCGGTTTGCGTGAGATGTACCCCGGACGTG---------CTCTGCCGAGCGATGCCAACT 1705
          |||            |||         :::          :::      |||   ||||||
Db    658 ProSerThrThrProCysProGlyThrIleThrIleThrIleValProAlaSerProThr 677

Qy   1706 TCACCATGCGTA------------------TGAGCTACGGCTCCA--------------- 1732
             |||||||||                  ||||||:::|||
Db    678 AspProCysValGluMetAspProSerThrGluAlaThrSerProProThrThrProLeu 697

Qy   1732 ------------------------------------------------------------ 1732

Db    698 ThrValPheProPheThrThrGluMetValThrCysProThrSerIleSerIleGlnThr 717

Qy   1732 ------------------------------------------------------------ 1732

Db    718 ThrLeuThrThrTyrMetAspThrSerSerMetMetProGluSerGluSerSerIleSer 737

Qy   1733 ---------TCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAG 1783
                   |||                     |||||||||      |||:::
Db    738 ProAsnAlaSerSerSerThrGlyThrGlyThrValProThrAsnThrValPheThrSer 757

Qy   1784 GCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATA 1843
             |||                                   ||||||
Db    758 ThrArg------------------------------------LeuPro---------Thr 762

Qy   1844 TCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACG---GTCAGCTCC 1900
             |||   |||        :::      :::::::::   :::|||      ||||||:::
Db    763 SerGluThrTrpLeuSerAsnSerSerValIleProLeuProLeuProGlyValSerThr 782

Qy   1901 ATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCG 1960
             |||   |||:::   :::   ::::::|||                     |||
Db    783 IleProLeuThrMetLysProSerSerSer------------------LeuProThrIle 796

Qy   1961 ATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTG 2020
             :::||||||:::      |||         |||      |||   :::
Db    797 LeuArgThrSerSerLysSerThrHisProSerProProThrThrArg------------ 812

Qy   2021 ACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCT 2080
                :::|||   |||:::   :::   |||            |||      |||   :::
Db    813 ---ThrSerGluThrProValAlaThrThrGln---------ThrProThrThrLeuThr 828

Qy   2081 TCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
             |||      |||      :::::::::   :::::::::::|||
Db    829 SerArgArgThrThrArgIleThrSerGlnMetThrThrGlnSer 843
```

RESULT 6
Q8TDH7

```
ID   Q8TDH7       PRELIMINARY;      PRT;  1029 AA.
AC   Q8TDH7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Intestinal membrane mucin MUC17 (Fragment).
GN   MUC17.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21845432; PubMed=11855812;
RA   Gum J.R. Jr., Crawley S.C., Hicks J.W., Szymkowski D.E., Kim Y.S.;
RT   "MUC17, a novel membrane-tethered mucin.";
RL   Biochem. Biophys. Res. Commun. 291:466-475(2002).
DR   EMBL; AF430017; AAL89737.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE   1029 AA;  108320 MW;  5E1AD19E4BF948E3 CRC64;

Alignment Scores:
Pred. No.:              0.0301        Length:        1029
Score:                  156.50        Matches:       147
Percent Similarity:     32.56%        Conservative:  63
Best Local Similarity:  22.79%        Mismatches:    230
Query Match:            4.10%         Indels:        205
DB:                     4             Gaps:          29

US-10-008-355-1 (1-2139) x Q8TDH7 (1-1029)

Qy    536 CTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTA 595
             :::|||   |||:::         :::||||||   :::||||||
```

```
Db     104 MetSerThrProSerGluGlySerSerSerLeuThrThrMet----------------- 117

Qy     596 TGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGT 655
                    ||||||   :::      |||:::|||          |||
Db     118 ---------LeuLeuSerSerThrTyrValThrSerSerGluAlaSerThrProSerThr 134

Qy     656 GGCCGCGTCACACGGGCGACT--------------------------------------- 676
              |||   |||:::|||
Db     135 ProSerValAspArgSerThrProValThrThrSerThrGlnSerAsnSerThrProThr 154

Qy     677 ------------------------------------------TCAGCGTATTCCGCG 691
                                                     |||
Db     155 ProProGluValIleThrLeuProMetSerThrProSerGluValSerThrProLeuThr 174

Qy     692 TGTATGCCGGTGCCGACAACC-----------GGCCGGCCGAATACAGCAAGGACAATA 739
            |||||||||   ||||||                              ||||||   |||:::
Db     175 IleMetProValSerThrThrSerValThrIleSerGluAlaGlyThrAlaSerThrLeu 194

Qy     740 ------------------------------------------------AACCCTATAAGC 751
                                                          :::|||::::::
Db     195 ProValAspThrSerThrProValIleThrSerThrGlnValSerSerSerProValThr 214

Qy     752 CCG---TTTACTTCGCTGCCGTAT------CCATGCAAG---GCTACAAGGCTGACGACT 799
           |||   |||:::::|||         |||   :::   :::|||   |||||||||
Db     215 ProGluGlyThrThrMetProIleTrpThrProSerGluGlySerThrProLeuThrThr 234

Qy     800 ATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGG 859
           ||||||   |||   :::|||                       |||   |||
Db     235 MetPro---ValSerThrThrArgValThrSerSerGluGlySerThrLeuSerThrPro 253

Qy     860 AAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCA 919
              :::|||:::   |||      :::      |||      |||:::||||||   |||
Db     254 SerValValThrSerThr---ProValThrThrSerThrGluAlaIleSerSerSerAla 272

Qy     920 TCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATG 979
           :::                          |||      :::   |||      ||||||
Db     273 Thr--------------------------LeuAspSerThrThrMetSerValSerMet 283

Qy     980 ---CTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTC 1036
              :::   :::   |||:::|||                              ||||||
Db     284 ProMetGluIleSerThrLeuGly--------------------ThrThrIleLeuVal 296

Qy    1037 TTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGA 1096
              |||      |||         |||                        ||||||
Db     297 SerThrThrProValThrArgPheProGluSerSerThrProSerIleProSerValTyr 316

Qy    1097 ACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAG 1156
           |||:::         :::||||||                   |||
Db     317 ThrSer------MetSerMetThrThrAlaSerGluGlySerSerSerProThrThrLeu 334

Qy    1157 GAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGG 1216
           |||               |||:::                     :::   ||||||:::::
Db     335 GluGlyThrThrThrMetProMetSerThrThrSerGluArgSerThrLeu--------- 351

Qy    1217 TGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTA 1276
                                                   |||      :::::::::   |||:::
Db     352 --------------------------------LeuThrThrValLeuIleSerProIle 360

Qy    1277 TCC-----------------------------------------------TCAAAT 1285
           |||                                               |||
Db     361 SerValMetSerProSerGluAlaSerThrLeuSerThrProProGlyAspThrSerThr 380

Qy    1286 CGCTTGACGACAAGTACAAAGACT------------------------------------ 1309
              |||   ||||||||||||
Db     381 ProLeuLeuThrSerThrLysAlaGlySerPheSerIleProAlaGluValThrThrIle 400

Qy    1310 ------------ACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTG 1357
                       ||||||   |||||||||                  |||   :::   |||
Db     401 ArgIleSerIleThrSerGluArgSerThr----------------ProLeuThrThrLeu 415

Qy    1358 TACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGA 1417
                       :::||||||||||||   |||
Db     416 LeuValSerThrThrLeuProThrSerPhePro-----------------------Gly 427

Qy    1418 AATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTT 1477
              |||   |||:::         :::   ||||||
Db     428 AlaSerIleAlaSerThrProProLeuAspThrSer----------------------- 439

Qy    1478 ATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTA 1537
                 |||:::   |||   :::         :::      ::::::|||
Db     440 ------ThrThrPheThrProSerThrAspThrAlaSerThrProThrIleProValAla 457

Qy    1538 TCG-----------AGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTG 1585
           :::               |||   :::   |||   |||
Db     458 ThrThrIleSerValSerValIleThrGluGlySerThrProGlyThr----------- 473

Qy    1586 CTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTC 1645
              :::|||   |||                     |||:::         |||   |||
Db     474 ---ThrIlePheIlePro---------SerThrProValThrSerSerThrAlaAspVal 489

Qy    1646 TTT--------------------TCTTTGCCGGTTTGCGTGAGATGT---------ACC 1675
           |||                    |||   ||||||                  |||
Db     490 PheProAlaThrThrGlyAlaValSerThrProValIleThrSerThrGluLeuAsnThr 509

Qy    1676 CCGGACGTGCTCTGCCGAGCGATGCCAACTTCA---------------------------- 1708
           |||             :::      ||||||
Db     510 ProSerThrSerSerSerSerThrThrThrSerPheSerThrThrLysGluPheThrThr 529

Qy    1709 CCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACA 1768
           |||   :::         ||||||               :::         |||   |||   |||
Db     530 ProAlaMetThrThrAlaAlaProLeuThrTyrValThrMetSerThrAlaProSerThr 549

Qy    1769 ACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTG 1828
                                                      :::   |||   |||   |||
Db     550 ProArgThrThrSerArgGlyCysThrThrSerAlaSerThrLeuSerAlaThrSerThr 569

Qy    1829 CCGTACAGGAGAATA--------------------------TCCTCGACCTCTTCCGCA 1861
           |||:::                                   ||||||   ||||||
Db     570 ProHisThrSerThrSerValThrThrArgProValThrProSerSerGluSerSerArg 589

Qy    1862 CCAAAAACTATGGTCGCT------ATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTAT 1915
           |||   |||:::   :::      :::|||   |||                  |||
Db     590 ProSerThrIleThrSerHisThrIleProProThrPheProProAlaHisSerSerThr 609

Qy    1916 CGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTC 1975
                 |||||||||   :::   |||   |||   |||   ::::::   |||
Db     610 ProProThrThrSerAlaSerSerThrThrValAsnProGluAlaValThrThrMetThr 629

Qy    1976 TGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAAC 2035
                          |||      |||               ||||||:::::::::
Db     630 ThrArgThrLysProSerThrArgThrThrSerPheProThrValThrThrThrAlaVal 649

Qy    2036 CCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAAT 2095
           |||                     |||:::   |||      |||            |||
Db     650 Pro-----------------------ThrAsnThrThrIleLysSer----------Asn 658

Qy    2096 GGGGTCAGTGCCCCC 2110
              |||   |||
Db     659 ProThrSerThrPro 663
```

RESULT 7
Q9UQ36

```
ID   Q9UQ36     PRELIMINARY;     PRT;   1275 AA.
AC   Q9UQ36;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RNA binding protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, partial cds.";
```

```
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016091; BAA83717.1; -.
FT   NON_TER      1       1
SQ   SEQUENCE   1275 AA;  136869 MW;  45C2B2F85E98A6F6 CRC64;

Alignment Scores:
Pred. No.:              0.0363        Length:          1275
Score:                  156.00        Matches:         184
Percent Similarity:     34.42%        Conservative:    81
Best Local Similarity:  23.90%        Mismatches:      276
Query Match:            4.08%         Indels:          229
DB:                     4             Gaps:            35

US-10-008-355-1 (1-2139) x Q9UQ36 (1-1275)

Qy     27 TCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGGGTAGCCAAAGCCGACAAAGGCATGTG 86
             |||   ||||||||   |||                 |||:::|||||         :::
Db    102 SerGlySerSerProGluVal-----------AspSerLysSerArg---------Leu 114

Qy     87 GCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCT 146
          :::|||:::|||::::::||||||   |||                          |||
Db    115 SerProArgArgSerArgSerGlySerSerProGluValLysAspLysProArgAlaAla 134

Qy    147 CCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTT 206
          |||
Db    135 Pro--------------------------------------------------------- 135

Qy    207 CGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCA 266
                                 :::||||||   |||
Db    136 -----------------------ArgAlaGlnSerGlySerAsp-------------- 142

Qy    267 CTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGG 326
                      |||||||||||         |||       |||
Db    143 -----------SerSerProGluProLysAlaProAlaProArgAlaLeuProArgArg 158

Qy    327 TTTCGT------TTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAA 380
             |||        |||      |||   |||   |||       ||||||
Db    159 SerArgSerGlySerSerSerLysGlyArgGlyProSerProGluGlySerSerSerThr 178

Qy    381 GTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCAC 440
          |||:::   :::                    |||||||||:::
Db    179 GluSerSerProGluHisProProLysSer---------ArgThrAlaArg--------- 192

Qy    441 TGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACT---GGCCAAAAA 497
             |||      |||:::                     ||||||         ::::::
Db    193 ---ArgGlySerArgSerSerProGluProLysThrLysSerArgThrProProArgArg 211

Qy    498 AGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAA------ 551
          |||      |||      |||   |||   |||:::|||   |||   ::::::
Db    212 ArgSerSerArgSerSerProGluLeuThrArgLysAlaArgLeuSerArgArgSerArg 231

Qy    552 ----------------------------CGAATACTTCCT---CATCGTCTACGATGT 578
                                      |||      |||   ||||||
Db    232 SerAlaSerSerSerProGluThrArgSerArgThrProProArgHisArg--------- 248

Qy    579 ATTCAAGGACGTTCGTATGGTATTTGCTCC---------TCCCAGCTCTGTAGGTAAGTT 629
                   ||||||   :::   |||           ::::::      |||
Db    249 ---------ArgSerProSerValSerSerProGluProAlaGluLysSerArgSerSer 265

Qy    630 CGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCG 689
          ||||||||                    :::||||||   |||   :::
Db    266 ArgArgArgArg---------------SerAlaSerSerProArgThrLysThrThrSer 280

Qy    690 CGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAA 749
          |||      |||         |||      |||
Db    281 ArgArgGlyArgSerProSerProLysProArg-------------------------- 291

Qy    750 GCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGAC 809
                |||   |||   |||      :::|||   :::
Db    292 ---GlyLeuGlnArgSerArgSerArgSerArgArgGluLys----------------- 304

Qy    810 CATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTAT 869
                                                              |||:::
Db    305 ---------------------------------------------ThrArgThrThr 308

Qy    870 CGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGA 929
          |||:::|||::::::|||            |||||||:::      :::|||      :::
Db    309 ArgArgArgAspArgSerGly------SerSerGlnSerThrSerArgArgArgGlnArg 326

Qy    930 AGCCATGAGCGCAGATCAGGCTACCCGTATCAAA--------------TATGCCAGCAA 974
          |||      |||      :::   ||||||   :::                |||   |||::
Db    327 SerArgSerArg-SerArgValThrArgArgArgGlyGlySerGlyTyrHisSerAr 346

Qy    975 GTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCG 1034
          :                             :::   |||               ||||||   :::||
Db    346 gSerProAlaArgGlnGluSerSerArgThrSerSerArgArgArgGlyArgSerAr 366

Qy   1035 TCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTC---GCAGACTGGATCCG 1091
          |                  |||||||||:::         |||         |||   |||   ||
Db    366 gThrProProThrSerArgLysArgSerArgSerArgThrSerProAlaProTrpLysAr 386

Qy   1092 TAAGAACGGCAAGAGTG-----------------CTGTCTATGGCGATGTATTGTCTTC 1133
          |        :::|||                          |||      :::         |||:
Db    386 gSerArg-SerArgAlaSerProAlaThrHisArgArgSerArgSerArgThrProLeuI 406

Qy   1134 TCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGA 1193
          ::|||:::|||      |||               |||      |||            :::|
Db    406 leSerArgArgArgSerArgSerArgThrSerProValSerArgArgArgSerArgSerA 426

Qy   1194 GA-----------------------------CGCTCTTCGGTGGTACCGAGGTGGTTCG 1223
          ||                                    |||:::|||   |||   |||
Db    426 rgThrSerValThrArgArgArgSerArgSerArgAlaSerProValSerArgArgArgS 446

Qy   1224 TTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAA 1283
                   |||                                        ||||||:::
Db    446 erArgSerArgThrPro-----------------------------ProValThrArgA 456

Qy   1284 ATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGC 1343
             |||         :::   |||   ||||||      ||||||      |||      ||||
Db    456 rgArgSerArgSerArgThrProThrThrArgArgArgSerArgSerArg---ThrProP 475

Qy   1344 CATGCTCGATATTGTACGCCGGCGTATCCCTGCCG------------------------ 1378
          ||   :::         :::         |||
Db    475 roValThrArgArgArgSerArgSerArgThrProProValThrArgArgArgSerArgS 495

Qy   1379 --ACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAA 1436
              :::|||||||||      |||         :::|||   |||      |||||||
Db    495 erArgThrSerProIleThrArgArgArgSerArgSerArgThrSerProValThrArgA 515

Qy   1437 GA------AGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTT 1490
          ||            :::||||||      ||||||                              |
Db    515 rgArgSerArgSerArgThrSerProValThrArg-----------------ArgArgS 529

Qy   1491 CCATGCCATGCTCAAGTCCATGGACAAGG--------------AAAAGTTTGCCAAGGC 1535
          ||         :::||||||   ||||||                     ::::::   |||   :
Db    529 erArgSerArgThrSerProValThrArgArgArgSerArgSerArgThrProProAlaI 549

Qy   1536 TATCGAGAAAGATCCGGCAGTAGAGC------TTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
          ::   |||:::   |||         :::         ||||||
Db    549 leArgArgArgSerArgSerArgThrProLeuLeuProArg------------------ 562

Qy   1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
                  :::         |||         |||:::   :::||||||   |||
Db    563 --------LysArgSerArgSerArgSerProLeuAlaIleArgArgArgSerArgSerA 580

Qy   1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCAC 1709
             |||         |||                     ||||::::::         ||||
Db    580 rgThrProArgThrAlaArgGlyLys---------------ArgSerLeuThrArgSerP 595

Qy   1710 CATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAA 1769
          ||                        |||:::      :::   ||||||:::         |||::::
Db    595 ro-----------------ProAla------IleArgArgArgSerAlaSerGlySerS 607

Qy   1770 CTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGC 1829
          ::      |||                                        |||:::      :::   |
Db    607 erSerAspArgSer-----------------------------ArgSerAlaThrProP 617
```

```
Qy   1830 CGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAA 1889
           ||        |||       |||    |||        |||                 |||
Db    617 roAlaThrArgAsnHisSerGlySerArgThrPro------------------ProValA 631

Qy   1890 CGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAG 1949
             ::::::|||       :::|||      :::        |||   :::   ||||||:::
Db    631 laLeuAsnSerSerArgMetSerCysPheSerArgProSerMetSerProThrProLeuA 651

Qy   1950 CCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGC 2009
                                                          :::   |||   |
Db    651 sp----------------------------------ArgCysArgSerProGlyMetL 659

Qy   2010 TATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGA------ 2063
           ||       :::   ||||||        |||:::    ||| |||   |||
Db    659 euGluProLeuGlySerSerArgThrProMetSerValLeuGln-GlnAlaGlyGlySer 678

Qy   2064 ---------CATCCGCTACGTTCTCTTCATGATTGACAAATGGGTCAGTGCCCCCGTCT 2114
                   |||       :::   |||   |||       |||||||||    :::
Db    679 MetMetAspGlyProGlyProArgIleProAspHisGlnArgThrSerValProGluAsn 698

Qy   2115 CATCCAAGAGCTGAAGTTGATCTA 2138
           |||   ::::::   :::   |||
Db    699 HisAlaGlnSerArgIleAlaLeu 706
```

RESULT 8
O15038
ID   O15038   PRELIMINARY;   PRT;   1783 AA.
AC   O15038;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA0324 protein (Fragment).
GN   KIAA0324.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
DR   EMBL; AB002322; BAA20782.2; -.
FT   NON_TER   1   1
SQ   SEQUENCE   1783 AA; 190940 MW; 660302F6FD4179AB CRC64;

Alignment Scores:

Pred. No.:	0.043	Length:	1783
Score:	156.00	Matches:	184
Percent Similarity:	34.42%	Conservative:	81
Best Local Similarity:	23.90%	Mismatches:	276
Query Match:	4.08%	Indels:	229
DB:	4	Gaps:	35

US-10-008-355-1 (1-2139) x O15038 (1-1783)

```
Qy     27 TCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGGGTAGCCAAAGCCGACAAAGGCATGTG 86
           |||    |||||||||   |||                 |||:::||||||      :::
Db    610 SerGlySerSerProGluVal------------AspSerLysSerArg----------Leu 622

Qy     87 GCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCT 146
           :::|||:::|||::::::||||||   |||                            |||
Db    623 SerProArgArgSerArgSerGlySerSerProGluValLysAspLysProArgAlaAla 642

Qy    147 CCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTT 206
           |||
Db    643 Pro--------------------------------------------------------- 643

Qy    207 CGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCA 266
                                   :::||||||   |||
Db    644 -----------------------ArgAlaGlnSerGlySerAsp--------------- 650

Qy    267 CTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGG 326
                                 |||||||||||          |||        |||
Db    651 ------------SerSerProGluProLysAlaProAlaProArgAlaLeuProArgArg 666

Qy    327 TTTCGT------TTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAA 380
              |||      |||    |||    |||   |||         ||||||
Db    667 SerArgSerGlySerSerSerLysGlyArgGlyProSerProGluGlySerSerSerThr 686

Qy    381 GTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCAC 440
              |||:::   :::                      |||||||||:::
Db    687 GluSerSerProGluHisProProLysSer---------ArgThrAlaArg--------- 700

Qy    441 TGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACT---GGCCAAAAA 497
              |||         |||:::                   ||||||      ::::::
Db    701 ---ArgGlySerArgSerSerProGluProLysThrLysSerArgThrProProArgArg 719

Qy    498 AGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAA------ 551
          |||    |||       |||  |||    |||:::|||   |||   ::::::
Db    720 ArgSerSerArgSerSerProGluLeuThrArgLysAlaArgLeuSerArgArgSerArg 739

Qy    552 ------------------------------CGAATACTTCCT---CATCGTCTACGATGT 578
                                       |||     |||   ||||||
Db    740 SerAlaSerSerSerProGluThrArgSerArgThrProProArgHisArg--------- 756

Qy    579 ATTCAAGGACGTTCGTATGGTATTTGCTCC---------TCCCAGCTCTGTAGGTAAGTT 629
                  ||||||    :::   |||          ::::::         |||
Db    757 ---------ArgSerProSerValSerSerProGluProAlaGluLysSerArgSerSer 773

Qy    630 CGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCG 689
          ||||||||||                :::||||||    |||     :::
Db    774 ArgArgArgArg---------------SerAlaSerSerProArgThrLysThrThrSer 788

Qy    690 CGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAA 749
          |||     |||         |||     |||
Db    789 ArgArgGlyArgSerProSerProLysProArg------------------------- 799

Qy    750 GCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGAC 809
              |||   |||    |||       :::|||    :::
Db    800 ---GlyLeuGlnArgSerArgSerArgSerArgArgGluLys------------------ 812

Qy    810 CATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTAT 869
                                                            |||:::
Db    813 ---------------------------------------------ThrArgThrThr 816

Qy    870 CGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGA 929
          |||:::|||:::::::|||        ||||||:::      :::|||     :::
Db    817 ArgArgArgAspArgSerGly------SerSerGlnSerThrSerArgArgArgGlnArg 834

Qy    930 AGCCATGAGCGCAGATCAGGCTACCCGTATCAAA---------------TATGCCAGCAA 974
          |||    |||    :::   ||||||   :::                  |||   |||::
Db    835 SerArgSerArg-SerArgValThrArgArgArgArgGlyGlySerGlyTyrHisSerAr 854

Qy    975 GTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCG 1034
          :                          :::   |||              ||||||   :::||
Db    854 gSerProAlaArgGlnGluSerSerArgThrSerSerArgArgArgGlyArgSerAr 874

Qy   1035 TCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTC---GCAGACTGGATCCG 1091
          |               |||||||||:::         |||          |||   |||  ||
Db    874 gThrProProThrSerArgLysArgSerArgSerArgThrSerProAlaProTrpLysAr 894

Qy   1092 TAAGAACGGCAAGAGTG-----------------CTGTCTATGGCGATGTATTGTCTTC 1133
          |          :::|||                 |||      :::          |||:
Db    894 gSerArg-SerArgAlaSerProAlaThrHisArgArgSerArgSerArgThrProLeuI 914

Qy   1134 TCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGA 1193
          ::|||:::|||       |||             |||      |||           :::|
Db    914 leSerArgArgArgSerArgSerArgThrSerProValSerArgArgArgSerArgSerA 934
```

```
Qy   1194 GA-----------------------------CGCTCTTCGGTGGTACCGAGGTGGTTCG 1223
          ||                             |||:::|||   |||   |||
Db    934 rgThrSerValThrArgArgArgSerArgSerArgAlaSerProValSerArgArgArgS 954

Qy   1224 TTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAA 1283
                   |||                             ||||||:::
Db    954 erArgSerArgThrPro-----------------------------ProValThrArgA 964

Qy   1284 ATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGC 1343
            |||       :::  |||  ||||||    ||||||    |||      ||||
Db    964 rgArgSerArgSerArgThrProThrThrArgArgArgSerArgSerArg---ThrProP 983

Qy   1344 CATGCTCGATATTGTACGCCGGCGTATCCCTGCCG------------------------- 1378
          ||  :::        :::        |||
Db    983 roValThrArgArgArgSerArgSerArgThrProProValThrArgArgArgSerArgS 1003

Qy   1379 --ACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAA 1436
                :::|||||||||       |||      :::|||   |||     |||||||
Db   1003 erArgThrSerProIleThrArgArgArgSerArgSerArgThrSerProValThrArgA 1023

Qy   1437 GA------AGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTT 1490
          ||         :::||||||   ||||||                            |
Db   1023 rgArgSerArgSerArgThrSerProValThrArg------------------ArgArgS 1037

Qy   1491 CCATGCCATGCTCAAGTCCATGGACAAGG--------------AAAAGTTTGCCAAGGC 1535
          ||        :::||||||   ||||||                :::::: |||   :
Db   1037 erArgSerArgThrSerProValThrArgArgArgSerArgSerArgThrProProAlaI 1057

Qy   1536 TATCGAGAAAGATCCGGCAGTAGAGC------TTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
          ::   |||:::   |||      :::          ||||||
Db   1057 leArgArgArgSerArgSerArgThrProLeuLeuProArg------------------ 1070

Qy   1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
                   :::       |||         |||:::   :::||||||   |||
Db   1071 --------LysArgSerArgSerArgSerProLeuAlaIleArgArgArgSerArgSerA 1088

Qy   1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCAC 1709
             |||         |||                         |||::::::   ||||
Db   1088 rgThrProArgThrAlaArgGlyLys--------------ArgSerLeuThrArgSerP 1103

Qy   1710 CATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAA 1769
          ||                  |||:::      :::  ||||||:::      |||:::::
Db   1103 ro------------------ProAla------IleArgArgArgSerAlaSerGlySerS 1115

Qy   1770 CTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGC 1829
          ::      |||                             |||:::    :::    |
Db   1115 erSerAspArgSer----------------------------ArgSerAlaThrProP 1125

Qy   1830 CGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAA 1889
          ||     |||      |||   |||      |||                    |||
Db   1125 roAlaThrArgAsnHisSerGlySerArgThrPro-------------------ProValA 1139

Qy   1890 CGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAG 1949
           ::::::|||       :::|||       :::      |||     :::   |||||:::
Db   1139 laLeuAsnSerSerArgMetSerCysPheSerArgProSerMetSerProThrProLeuA 1159

Qy   1950 CCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGC 2009
                                                       :::   |||   |
Db   1159 sp----------------------------------ArgCysArgSerProGlyMetL 1167

Qy   2010 TATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGA------ 2063
          ||      :::   ||||||      |||:::          ||| |||   |||
Db   1167 euGluProLeuGlySerSerArgThrProMetSerValLeuGln-GlnAlaGlyGlySer 1186

Qy   2064 --------CATCCGCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCT 2114
                  |||         :::   |||   |||       |||||||||    :::
Db   1187 MetMetAspGlyProGlyProArgIleProAspHisGlnArgThrSerValProGluAsn 1206

Qy   2115 CATCCAAGAGCTGAAGTTGATCTA 2138
          |||   ::::::   :::   |||
Db   1207 HisAlaGlnSerArgIleAlaLeu 1214
```

RESULT 9

```
O60382
ID   O60382      PRELIMINARY;      PRT;  1791 AA.
AC   O60382;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA0324 (Fragment).
GN   KIAA0324.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA   Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA   Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,
RA   Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RT   "Sequencing of Human Chromosome 16p13.3.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ricke D.O.;
RT   "Large Scale Sequence Analysis and Annotation with the Sequence
RT   Comparison Analysis (SCAN) System.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004493; AAC08453.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   1791 AA;  191306 MW;  3A7B5530AEE95F3E CRC64;

Alignment Scores:
Pred. No.:               0.0431        Length:        1791
Score:                   156.00        Matches:       184
Percent Similarity:      34.42%        Conservative:  81
Best Local Similarity:   23.90%        Mismatches:    276
Query Match:             4.08%         Indels:        229
DB:                      4             Gaps:          35

US-10-008-355-1 (1-2139) x O60382 (1-1791)

Qy     27 TCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGGGTAGCCAAAGCCGACAAAGGCATGTG 86
           |||   |||||||||   |||                |||:::||||||        :::
Db    611 SerGlySerSerProGluVal------------AspSerLysSerArg---------Leu 623

Qy     87 GCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCT 146
           :::|||:::|||::::::||||||   |||                              |||
Db    624 SerProArgArgSerArgSerGlySerSerProGluValLysAspLysProArgAlaAla 643

Qy    147 CCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTT 206
           |||
Db    644 Pro--------------------------------------------------------- 644

Qy    207 CGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCA 266
                                   :::||||||   |||
Db    645 -----------------------ArgAlaGlnSerGlySerAsp-------------- 651

Qy    267 CTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGG 326
                      ||||||||||||                |||         |||
Db    652 -----------SerSerProGluProLysAlaProAlaProArgAlaLeuProArgArg 667

Qy    327 TTTCGT------TTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAA 380
             |||          |||      |||    |||    |||          ||||||
Db    668 SerArgSerGlySerSerSerLysGlyArgGlyProSerProGluGlySerSerSerThr 687

Qy    381 GTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCAC 440
             |||:::   :::                          |||||||||:::
Db    688 GluSerSerProGluHisProProLysSer---------ArgThrAlaArg--------- 701

Qy    441 TGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACT---GGCCAAAAA 497
             |||             |||:::                    ||||||       ::::::
Db    702 ---ArgGlySerArgSerSerProGluProLysThrLysSerArgThrProProArgArg 720

Qy    498 AGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAA------ 551
```

```
                |||      |||      |||    |||    |||:::|||    |||     ::::::
Db      721 ArgSerSerArgSerSerProGluLeuThrArgLysAlaArgLeuSerArgArgSerArg 740

Qy      552 -----------------------------CGAATACTTCCT---CATCGTCTACGATGT 578
                                             |||      |||    ||||||
Db      741 SerAlaSerSerSerProGluThrArgSerArgThrProProArgHisArg--------- 757

Qy      579 ATTCAAGGACGTTCGTATGGTATTTGCTCC---------TCCCAGCTCTGTAGGTAAGTT 629
                         ||||||     :::    |||          ::::::     |||
Db      758 ---------ArgSerProSerValSerSerProGluProAlaGluLysSerArgSerSer 774

Qy      630 CGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCG 689
            ||||||||                      :::||||||      |||     :::
Db      775 ArgArgArgArg--------------SerAlaSerSerProArgThrLysThrThrSer 789

Qy      690 CGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAA 749
            |||      |||         |||      |||
Db      790 ArgArgGlyArgSerProSerProLysProArg-------------------------- 800

Qy      750 GCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGAC 809
                  |||    |||    |||        :::|||     :::
Db      801 ---GlyLeuGlnArgSerArgSerArgSerArgArgGluLys----------------- 813

Qy      810 CATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTAT 869
                                                              |||:::
Db      814 ------------------------------------------------ThrArgThrThr 817

Qy      870 CGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGA 929
            |||:::|||::::::|||             ||||||:::      :::|||      :::
Db      818 ArgArgArgAspArgSerGly------SerSerGlnSerThrSerArgArgArgGlnArg 835

Qy      930 AGCCATGAGCGCAGATCAGGCTACCCGTATCAAA---------------TATGCCAGCAA 974
            |||      |||     :::    ||||||    :::              |||   |||::
Db      836 SerArgSerArg-SerArgValThrArgArgArgArgGlyGlySerGlyTyrHisSerAr 855

Qy      975 GTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCG 1034
            :                       :::    |||              ||||||    :::||
Db      855 gSerProAlaArgGlnGluSerSerArgThrSerSerArgArgArgArgGlyArgSerAr 875

Qy     1035 TCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTC---GCAGACTGGATCCG 1091
            |                  ||||||||||:::      |||           |||   |||   ||
Db      875 gThrProProThrSerArgLysArgSerArgSerArgThrSerProAlaProTrpLysAr 895

Qy     1092 TAAGAACGGCAAGAGTG-----------------CTGTCTATGGCGATGTATTGTCTTC 1133
            |        :::|||                           |||    :::        |||:
Db      895 gSerArg-SerArgAlaSerProAlaThrHisArgArgSerArgSerArgThrProLeuI 915

Qy     1134 TCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGA 1193
            ::|||:::|||       |||            |||       |||             :::|
Db      915 leSerArgArgArgSerArgSerArgThrSerProValSerArgArgArgSerArgSerA 935

Qy     1194 GA----------------------------CGCTCTTCGGTGGTACCGAGGTGGTTCG 1223
            ||                               |||:::|||    |||    |||
Db      935 rgThrSerValThrArgArgArgSerArgSerArgAlaSerProValSerArgArgArgS 955

Qy     1224 TTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAA 1283
                           |||                              ||||||:::
Db      955 erArgSerArgThrPro----------------------------ProValThrArgA 965

Qy     1284 ATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGC 1343
              |||      :::    |||   ||||||        ||||||     |||       ||||
Db      965 rgArgSerArgSerArgThrProThrThrArgArgArgSerArgSerArg---ThrProP 984

Qy     1344 CATGCTCGATATTGTACGCCGGCGTATCCCTGCCG------------------------ 1378
            ||   :::            :::          |||
Db      984 roValThrArgArgArgSerArgSerArgThrProProValThrArgArgArgSerArgS 1004

Qy     1379 --ACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAA 1436
                 :::|||||||||        |||          :::|||    |||        ||||||
Db     1004 erArgThrSerProIleThrArgArgArgSerArgSerArgThrSerProValThrArgA 1024

Qy     1437 GA------AGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTT 1490
            ||                 :::||||||       ||||||                  |
Db     1024 rgArgSerArgSerArgThrSerProValThrArg-----------------ArgArgS 1038

Qy     1491 CCATGCCATGCTCAAGTCCATGGACAAGG---------------AAAAGTTTGCCAAGGC 1535
            ||          :::||||||    ||||||                 ::::::   |||   :
Db     1038 erArgSerArgThrSerProValThrArgArgSerArgSerArgThrProProAlaI 1058

Qy     1536 TATCGAGAAAGATCCGGCAGTAGAGC------TTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
            ::   |||:::    |||          :::          ||||||
Db     1058 leArgArgArgSerArgSerArgThrProLeuLeuProArg------------------ 1071

Qy     1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
                          :::       |||         |||:::     :::||||||    |||
Db     1072 --------LysArgSerArgSerArgSerProLeuAlaIleArgArgArgSerArgSerA 1089

Qy     1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCAC 1709
                   |||          |||                        |||::::::      ||||
Db     1089 rgThrProArgThrAlaArgGlyLys--------------ArgSerLeuThrArgSerP 1104

Qy     1710 CATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAA 1769
            ||                     |||:::        :::    ||||||:::        |||::::
Db     1104 ro-----------------ProAla------IleArgArgArgSerAlaSerGlySerS 1116

Qy     1770 CTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGC 1829
            ::       |||                                       |||:::    :::    |
Db     1116 erSerAspArgSer-----------------------------ArgSerAlaThrProP 1126

Qy     1830 CGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAA 1889
            ||      |||       |||     |||      |||                      |||
Db     1126 roAlaThrArgAsnHisSerGlySerArgThrPro-----------------ProValA 1140

Qy     1890 CGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAG 1949
              ::::::|||      :::|||       :::          |||    :::    ||||||:::
Db     1140 laLeuAsnSerSerArgMetSerCysPheSerArgProSerMetSerProThrProLeuA 1160

Qy     1950 CCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGC 2009
                                                                  :::    |||    |
Db     1160 sp---------------------------------ArgCysArgSerProGlyMetL 1168

Qy     2010 TATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGA------ 2063
              ||        :::    ||||||        |||:::            |||   |||     |||
Db     1168 euGluProLeuGlySerSerArgThrProMetSerValLeuGln-GlnAlaGlyGlySer 1187

Qy     2064 ---------CATCCGCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCT 2114
                            |||          :::    |||     |||           |||||||||    :::
Db     1188 MetMetAspGlyProGlyProArgIleProAspHisGlnArgThrSerValProGluAsn 1207

Qy     2115 CATCCAAGAGCTGAAGTTGATCTA 2138
              |||    ::::::    :::    |||
Db     1208 HisAlaGlnSerArgIleAlaLeu 1215

RESULT 10
Q9UQ35
ID   Q9UQ35      PRELIMINARY;      PRT;   2752 AA.
AC   Q9UQ35;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RNA binding protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, complete cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016092; BAA83718.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   2752 AA;   299672 MW;   109C64F181097123 CRC64;

Alignment Scores:
```

```
Pred. No.:              0.0535     Length:        2752
Score:                  156.00     Matches:       184
Percent Similarity:     34.42%     Conservative:  81
Best Local Similarity:  23.90%     Mismatches:    276
Query Match:            4.08%      Indels:        229
DB:                     4          Gaps:          35

US-10-008-355-1 (1-2139) x Q9UQ35 (1-2752)

Qy     27 TCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGGGTAGCCAAAGCCGACAAAGGCATGTG 86
Db   1579 SerGlySerSerProGluVal-----------AspSerLysSerArg---------Leu 1591

Qy     87 GCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCT 146
Db   1592 SerProArgArgSerArgSerGlySerSerProGluValLysAspLysProArgAlaAla 1611

Qy    147 CCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTT 206
Db   1612 Pro--------------------------------------------------------- 1612

Qy    207 CGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCA 266
Db   1613 -----------------------ArgAlaGlnSerGlySerAsp-------------- 1619

Qy    267 CTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGG 326
Db   1620 -----------SerSerProGluProLysAlaProAlaProArgAlaLeuProArgArg 1635

Qy    327 TTTCGT------TTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAA 380
Db   1636 SerArgSerGlySerSerSerLysGlyArgGlyProSerProGluGlySerSerSerThr 1655

Qy    381 GTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCAC 440
Db   1656 GluSerSerProGluHisProProLysSer---------ArgThrAlaArg--------- 1669

Qy    441 TGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACT---GGCCAAAAA 497
Db   1670 ---ArgGlySerArgSerSerProGluProLysThrLysSerArgThrProProArgArg 1688

Qy    498 AGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAA------ 551
Db   1689 ArgSerSerArgSerSerProGluLeuThrArgLysAlaArgLeuSerArgArgSerArg 1708

Qy    552 -----------------------------CGAATACTTCCT---CATCGTCTACGATGT 578
Db   1709 SerAlaSerSerSerProGluThrArgSerArgThrProProArgHisArg--------- 1725

Qy    579 ATTCAAGGACGTTCGTATGGTATTTGCTCC---------TCCCAGCTCTGTAGGTAAGTT 629
Db   1726 ---------ArgSerProSerValSerSerProGluProAlaGluLysSerArgSerSer 1742

Qy    630 CGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCG 689
Db   1743 ArgArgArgArg--------------SerAlaSerSerProArgThrLysThrThrSer 1757

Qy    690 CGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAA 749
Db   1758 ArgArgGlyArgSerProSerProLysProArg------------------------- 1768

Qy    750 GCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGAC 809
Db   1769 ---GlyLeuGlnArgSerArgSerArgSerArgArgGluLys----------------- 1781

Qy    810 CATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTAT 869
Db   1782 ----------------------------------------------ThrArgThrThr 1785

Qy    870 CGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGA 929
Db   1786 ArgArgArgAspArgSerGly------SerSerGlnSerThrSerArgArgArgGlnArg 1803

Qy    930 AGCCATGAGCGCAGATCAGGCTACCCGTATCAAA--------------TATGCCAGCAA 974
Db   1804 SerArgSerArg-SerArgValThrArgArgArgArgGlyGlySerGlyTyrHisSerAr 1823

Qy    975 GTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCG 1034
Db   1823 gSerProAlaArgGlnGluSerSerArgThrSerSerArgArgArgArgGlyArgSerAr 1843

Qy   1035 TCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTC---GCAGACTGGATCCG 1091
Db   1843 gThrProProThrSerArgLysArgSerArgSerArgThrSerProAlaProTrpLysAr 1863

Qy   1092 TAAGAACGGCAAGAGTG-----------------CTGTCTATGGCGATGTATTGTCTTC 1133
Db   1863 gSerArg-SerArgAlaSerProAlaThrHisArgArgSerArgSerArgThrProLeuI 1883

Qy   1134 TCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGA 1193
Db   1883 leSerArgArgArgSerArgSerArgThrSerProValSerArgArgArgSerArgSerA 1903

Qy   1194 GA----------------------------CGCTCTTCGGTGGTACCGAGGTGGTTCG 1223
Db   1903 rgThrSerValThrArgArgArgSerArgSerArgAlaSerProValSerArgArgArgS 1923

Qy   1224 TTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAA 1283
Db   1923 erArgSerArgThrPro----------------------------ProValThrArgA 1933

Qy   1284 ATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGC 1343
Db   1933 rgArgSerArgSerArgThrProThrThrArgArgArgSerArgSerArg---ThrProP 1952

Qy   1344 CATGCTCGATATTGTACGCCGGCGTATCCCTGCCG------------------------ 1378
Db   1952 roValThrArgArgArgSerArgSerArgThrProProValThrArgArgArgSerArgS 1972

Qy   1379 --ACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAA 1436
Db   1972 erArgThrSerProIleThrArgArgArgSerArgSerArgThrSerProValThrArgA 1992

Qy   1437 GA------AGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTT 1490
Db   1992 rgArgSerArgSerArgThrSerProValThrArg-----------------ArgArgS 2006

Qy   1491 CCATGCCATGCTCAAGTCCATGGACAAGG--------------AAAAGTTTGCCAAGGC 1535
Db   2006 erArgSerArgThrSerProValThrArgArgArgSerArgSerArgThrProProAlaI 2026

Qy   1536 TATCGAGAAAGATCCGGCAGTAGAGC------TTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
Db   2026 leArgArgArgSerArgSerArgThrProLeuLeuProArg------------------ 2039

Qy   1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
Db   2040 --------LysArgSerArgSerArgSerProLeuAlaIleArgArgArgSerArgSerA 2057

Qy   1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCAC 1709
Db   2057 rgThrProArgThrAlaArgGlyLys--------------ArgSerLeuThrArgSerP 2072

Qy   1710 CATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAA 1769
Db   2072 ro-----------------ProAla------IleArgArgArgSerAlaSerGlySerS 2084

Qy   1770 CTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGC 1829
Db   2084 erSerAspArgSer-----------------------------ArgSerAlaThrProP 2094

Qy   1830 CGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAA 1889
Db   2094 roAlaThrArgAsnHisSerGlySerArgThrPro-----------------ProValA 2108

Qy   1890 CGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAG 1949
```

```
               :::::|||        :::|||       :::        |||    :::    |||||||:::
Db    2108 laLeuAsnSerSerArgMetSerCysPheSerArgProSerMetSerProThrProLeuA 2128

Qy    1950 CCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGC 2009
                                                             :::   |||   |
Db    2128 sp---------------------------------ArgCysArgSerProGlyMetL 2136

Qy    2010 TATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGA------ 2063
            ||       :::   ||||||         |||:::          |||  |||     |||
Db    2136 euGluProLeuGlySerSerArgThrProMetSerValLeuGln-GlnAlaGlyGlySer 2155

Qy    2064 ---------CATCCGCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCT 2114
                       |||          :::   |||   |||      ||||||||    :::
Db    2156 MetMetAspGlyProGlyProArgIleProAspHisGlnArgThrSerValProGluAsn 2175

Qy    2115 CATCCAAGAGCTGAAGTTGATCTA 2138
              |||   ::::::   :::   |||
Db    2176 HisAlaGlnSerArgIleAlaLeu 2183
```

RESULT 11
Q9C105

```
ID   Q9C105      PRELIMINARY;      PRT;  1236 AA.
AC   Q9C105;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative glucoamylase I (alpha-1,4-glucan glucosidase), extracellular
DE   starch-degrading enzyme, by similarity to S. cerevisiae STA1, contains
DE   chitinase family signature.
GN   SPAPB1E7.04C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Wood V., Rajandream M.A., Barrell B.G., Seeger K., Harris D.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; AL590605; CAC36921.1; -.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   InterPro; IPR001917; NHtransf_2.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   PROSITE; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
KW   Glycosidase; Hydrolase.
SQ   SEQUENCE   1236 AA;  123387 MW;  5A2D33A30B87CDD8 CRC64;

Alignment Scores:
Pred. No.:              0.0455       Length:         1236
Score:                  154.50       Matches:        145
Percent Similarity:     35.92%       Conservative:   110
Best Local Similarity:  20.42%       Mismatches:     258
Query Match:            4.04%        Indels:         198
DB:                     3            Gaps:           24

US-10-008-355-1 (1-2139) x Q9C105 (1-1236)

Qy     166 AGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGT 225
            |||||||||        ||||||                             |||   ||||||:::|||
Db     215 SerPheAspPheLeuSerIleHisThrPheAsnSerSerThrGlyGluGlyCysSerGly 234

Qy     226 --------------------------ATCACAGTGTCCGATCAGGGCCTGATCTTTACC 258
                                     :::        ::::::                |||
Db     235 SerArgAsnSerThrPheAspAlaTrpValGluTyrAlaGluAspSerAlaTyrAsnThr 254

Qy     259 AACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTG 318
            |||            ||||||                                  |||   |||
Db     255 AsnThrSerLeuPheTyrGly------------------ValValGlyHisGlnAsnGly 268

Qy     319 CGCGATGGTTTCGTTTCT---CGCACGATGGGTGAGGAGCTT------------------ 357
                :::|||||||:::|||    :::    :::         :::|||
Db     269 SerAsnGlyPheIleSerProLysAsnLeuThrArgAspLeuLeuAsnTyrLysAlaAsn 288

Qy     358 ----------------------------------CCGATTCCGGGTCTTTCC------ 375
                                                :::        :::|||
Db     289 SerThrLeuPheGlyGlyValThrIleTrpAspThrSerLeuAlaAlaMetSerTyrAsp 308

Qy     376 -----------------GTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTA 417
                              |||:::   :::   ||||||:::               |||
Db     309 AsnSerSerGluThrPheValGluAlaIleHisLysIleLeuAspThrLysSerLys--- 327

Qy     418 GAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTA 477
                                |||       |||       :::   |||                 :::
Db     328 HisSerSerSerLysSerSerHisAspSerSerGlnGlyLeuGluSerThrSerSerIle 347

Qy     478 TGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCT 537
             |||   |||   |||              :::    :::    ::::::|||:::|||    |||
Db     348 -AlaLeuAsn---ProThrSerSerIleSerSerThrSerSerSerSerSerThrSerSe 366

Qy     538 TTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATG 597
             :::|||    |||     :::                :::    |||:::     |||
Db     366 rAlaIleSerThrIleSerGlnAspHisThrLysThrValThrSer-------------- 381

Qy     598 GTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGG 657
                                                  ||||||    |||          |||      ||
Db     382 ------------------------ValSerAspGluProThrThrIleThrAlaSerGl 393

Qy     658 CCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCG 717
             |      |||       |||:::                            ||||||
Db     393 y---AlaThrSerValThrThrThrThrLysThrAspPheAspThrValThrThrThrIl 412

Qy     718 GCCGAATACAGCAAGGACAATAAACCCT--------------------ATAAGCCCGTT 756
                  :::|||:::          |||:::                      :::|||   ::
Db     412 eValSerThrSerThrLeuIleSerAlaSerAspSerThrSerIleIleValSerSerTy 432

Qy     757 TACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGT 816
             :   |||                         :::||||||:::    |||              |||::
Db     432 rValSerThrValThrGlnProAlaSerThrArgValGlnThrThrThrValSerSerIl 452

Qy     817 TTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAAC 876
             :|||      :::|||:::                                              ||
Db     452 eSerThrSerValLys---------------------------------------GlnProTh 460

Qy     877 GAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATG 936
             |    ::::::          |||        :::|||    |||      :::|||    :::
Db     460 rAlaSerValAlaSerSerSerValSerValProSerSerSerSerValGln-------- 477

Qy     937 AGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTAT 996
                                |||    |||:::    |||    |||
Db     478 ---------------ProGlnSerSerThrProIleSer------------------- 485

Qy     997 TGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAG 1056
                                                                       ||
Db     486 --------------------------------------------------------Se 486

Qy    1057 CGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTAT 1116
             |                  |||:::              ::::::    |||:::|||
Db     486 rSerSerSerAlaSerSerProGlnSerThrLeuSerThrSerSerGluValValSerGl 506

Qy    1117 GGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAG 1176
                  |||||||||
Db     506 uValSerSerThrLeuLeuSer----------------------------------------- 513

Qy    1177 ATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTT 1236
                                     ||||||    :::|||
Db     514 ------------------GlySerSerAlaIlePro---------------------- 519

Qy    1237 GCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGAC 1296

Db     519 ------------------------------------------------------------ 519

Qy    1297 AAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATT 1356
             ||||||    :::::::::|||    |||::::::              |||        |||:::||
```

```
Db    520 -SerThrSerSerSerThrProSerSerSerIleIleSerSerProMetThrSerValLe 539

Qy   1357 GTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATA---TATTCAAGAATGTAATCGAC 1413
          |  :::      |||:::||||||||||||          :::|||          ::::||
Db    539 uSerSerSerSerSerIleProThrSerSerSerSerAspPheSerSerSerIleThrTh 559

Qy   1414 AAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTT 1473
          |  :::|||       :::       |||:::   ::::::::|||:::
Db    559 rIleSerSerGlyIleSerSerSerSerIleProSerThrPheSerSerValSerSerIl 579

Qy   1474 CCTTATAGCGACAAGTTCCATGCCATGCTCAAG------------------TCCATGGA 1514
           |||   ::::::::|||    |||   |||                       |||
Db    579 eLeuSerSerSerThrSerSerProSerSerThrSerLeuSerIleSerSerSerSerTh 599

Qy   1515 CAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGT 1574
                          |||:::   |||        |||                 |||
Db    599 rSerSerThrPheSerSerAlaSerThrSerSerPro-----------------SerSe 613

Qy   1575 AATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAA 1634
           |||:::::::::   :::     :::      ::::::::: |||   |||
Db    613 rIleSerSerSerIleSerSerSerSerThrIleLeuSerSer-ProThrProSerThrS 633

Qy   1635 GGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAG 1694
             :::   :::   |||               :::     :::     :::|||      :
Db    633 erSerLeuMetIleSerSerSerSerIleIleSerGlySerSerSerIleLeuSerSerS 653

Qy   1695 CGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGA 1754
          :::::  |||  |||               ::::::             :::
Db    653 erIleSerThrIlePro---------IleSerSerSerLeuSerThrTyrSerSerSerV 670

Qy   1755 CGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAA 1814
             :::|||   ::::::   :::      :::                |||      |||
Db    670 alIleProSerSerSerThrLeuValSerSer--------------SerSerSerLeuI 685

Qy   1815 GAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGG 1874
                      |||   |||         ::::::|||:::||||||:::|||
Db    685 leValSerSerSerPro---------ValAlaSerSerSerSerSerProIleProSerS 702

Qy   1875 TCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGG 1934
             :::          :::      |||   |||||||||       |||   :::|||    :
Db    702 erSerSerLeuValSerThrTyrSerAlaSerLeuSerAsnIleThrHisSerSerLeuS 722

Qy   1935 CGGTAACTCCGGTAGCCCCGTATTCG---ATAAGAACGGCCGTCTGATCGGTCTTGCTTT 1991
          ::::::|||   ::::::         |||   |||   |||:::|||   |||      |||
Db    722 erLeuThrAlaMetSerSerSerSerAlaIleProThrSerValAsnSerSerThrLeu- 741

Qy   1992 CGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAAC-----------CCGA 2039
                                :::        |||||||||||              |
Db    742 ---------------------IleThrAlaSerSerSerAsnThrLeuLeuSerSerI 754

Qy   2040 TCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGG 2099
           ||   |||:::     :::      :::|||   |||   |||:::        :::|||
Db    754 leThrSerSerSerAlaIleValSerSerThrThrValSerAsnIleSerSerAsnLeuP 774

Qy   2100 TCAGTGCCCCCGTCTCATCCAAGAGC 2125
             ||||||      ||||||:::|||
Db    774 roSerAlaThrAlaSerSerGlnSer 782
```

RESULT 12
Q8WWQ5

```
ID   Q8WWQ5      PRELIMINARY;      PRT;  2448 AA.
AC   Q8WWQ5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Mucin 5 (Fragment).
GN   MUC5AC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TRACHEA;
RX   MEDLINE=21426417; PubMed=11535137;
RA   Escande F., Aubert J.P., Porchet N., Buisine M.P.;
RT   "Human mucin gene MUC5AC: organization of its 5'-region and central
RT   repetitive region.";
RL   Biochem. J. 358:763-772(2001).
DR   EMBL; AJ298317; CAC83674.1; -.
DR   InterPro; IPR001209; Ribosomal_S14.
DR   InterPro; IPR002919; TIL_Cysrich.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR001846; VWF_D.
DR   Pfam; PF01826; TIL; 2.
DR   Pfam; PF00094; vwd; 3.
DR   SMART; SM00214; VWC; 3.
DR   SMART; SM00216; VWD; 3.
DR   PROSITE; PS00527; RIBOSOMAL_S14; UNKNOWN_2.
FT   NON_TER    2448   2448
SQ   SEQUENCE   2448 AA;  260970 MW;  A5C1BD627844D952 CRC64;

Alignment Scores:
Pred. No.:                0.113        Length:          2448
Score:                    151.00       Matches:         188
Percent Similarity:       35.64%       Conservative:    100
Best Local Similarity:    23.27%       Mismatches:      308
Query Match:              3.95%        Indels:          213
DB:                       4            Gaps:            41

US-10-008-355-1 (1-2139) x Q8WWQ5 (1-2448)

Qy     46 TTGGGTGCTTCAGGGGTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAA------ 99
           ||||||          :::           :::|||:::     |||||||:::
Db   1637 LeuGlyGlnAspValIleCysSerHisThrGluGlyLeuIleCysLeuAsnLysAsnGln 1656

Qy    100 ---------------CTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACG 144
                          |||   |||  |||   |||:::|||           :::
Db   1657 LeuProProIleCysTyrAsnTyrGlu-IleArgIleGlnCysCysGluThrValAsnVa 1676

Qy    145 CTCCCGTTGGATTCGC---------TCTACAGTTTCGACAAGCCGTCCATTGCCAATGCC 195
                |||   |||              ||||||:::|||:::||||||   |||   ||
Db   1676 lCysArgAspIleThrArgLeuProLysThrValAlaThrThrArgProThrProHisPr 1696

Qy    196 GTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTT 255
           |      :::                    :::         |||         :::
Db   1696 oThrGlyAlaGlnThrGlnThrThrPheThrThrHisMetProSerAlaSerThrGluGl 1716

Qy    256 ACCAACCACC------------------ACTGCGGATACGGTGCTATCCAGAGCCAAAGC 297
           ||||||                     ||||||   :::|||
Db   1716 nProThrAlaThrSerArgGlyGlyProThrAlaThrSerValThrGlnGlyThrHisTh 1736

Qy    298 ACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAG----GA 353
              :::|||          |||                   :::|||          ::
Db   1736 rThrLeuValThrArgAsnCysHisProArgCysThrTrpThrLysTrpPheAspValAs 1756

Qy    354 GCTTCCGATTCCGGGTCTT---------TCCGTGAAGTATCTGCGCAAGATCGTGAAGGT 404
            :   |||   ||||||                     :::            |||::::::
Db   1756 pPheProSerProGlyProHisGlyGlyAspLysGluThrTyrAsnAsnIleIleArgSe 1776

Qy    405 AACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAA 464
              :::|||:::      :::   :::   ||||||            |||   :::
Db   1776 rGlyGluLysIleCysArgArgProGluGluIleThrArgValGlnCysArgAlaLysSe 1796

Qy    465 AGCTCAGGAGGTATGCCAAGAA--------------------CTGGCCAAAAAAGAA-- 501
              ||||||   |||                                  ::::::::|||
Db   1796 rHisProGluValSerIleGluHisLeuGlyGlnValValGlnCysSerArgGluGl 1816

Qy    502 -------------AATGCAGACGAGAAC---------CAACTCTGCATCGTAGAGCCTTT 539
                       |||   |||:::              ::::::|||:::
Db   1816 yLeuValCysArgAsnGlnAspGlnGlnGlyProPheLysMetCysLeu---------- 1832

Qy    540 CTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGT 599
                  |||   |||      ::::::      :::      :::       |||
Db   1833 -------AsnTyrGluValArgValLeuCysCysGluThrProArgGlyCysHisMetTh 1850
```

```
Qy     600 ATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCC 659
               :::   |||  |||      ||||||   |||     |||
Db    1850 r---SerThrProGlySerThrSer-SerSerProAlaGlnThrThrProSerThrThrS 1869

Qy     660 GCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGC 719
                 |||           |||     :::     |||     |||   |||
Db    1869 erLysThrThrGluIleGlnAlaSerGlySerSerAlaProSerSerThrProGlyThrV 1889

Qy     720 CG------AATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATC 773
                   :::|||||||||||                   ||||||:::
Db    1889 alSerLeuSerThrAlaArgThrThrProAlaProGlyThrAlaThrSerVal------- 1906

Qy     774 CATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGA 833
                |||   |||     |||     |||   |||  |||
Db    1907 -----LysLysThrPheSerThrProSerPro---ProProValPro------------- 1919

Qy     834 TCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTAT 893
             |||||||||                      |||       |||
Db    1920 --AlaThrSer--------------------ThrSerSerMetSerThrThrAlaProG 1932

Qy     894 CGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTAC 953
                     |||||||||||   :::     :::|||
Db    1932 lyThrSerValValSerSerLysProThrProThrGluPro------------------S 1946

Qy     954 CCGTATCAAATATGC---CAGCAAGTATGCTCAGAGTGCTAACTATTGGAA--------- 1001
                |||::: |||  |||::::::|||:::          ::::::
Db    1946 erThrSerSer-CysLeuGlnGluLeuCysThrTrpThrGluTrpIleAspGlySerTyr 1965

Qy    1002 -----------------------GAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCT 1037
                                 :::||||||   ::::::             :::
Db    1966 ProAlaProGlyIleAsnGlyGlyAspPheAspThrPheGln---------------Asn 1980

Qy    1038 TGACGTGAT-------------------------AGGTCGTAAGCGTGCCGAGGAAAG 1070
              ||||||                        ||||||      ||||||  :::
Db    1981 LeuArgAspGluGlyTyrThrPheCysGluSerProArgSerValGlnCysArgAlaGlu 2000

Qy    1071 AGCATTCGCAGACTGG---ATCCGTAAGAACGGCAAGAGTGCTGTC------TATGGCGA 1121
           ||| |||    :::    ::: :::    |||:::     :::      :::  ::
Db    2001 Ser-PheProAsnThrProLeuGlyArgLeuGlyGlnAspValIleCysSerHisThrGl 2020

Qy    1122 TGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATG-- 1179
           :    |||     |||   |||                      |||  |||:::
Db    2020 uGlyLeuIleCysLeuAsnLysAsnGlnLeuProProIleCysTyrAsnTyrGluIleAr 2040

Qy    1180 -ACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGC 1238
                       ||||||:::           ::::::  |||     :::
Db    2040 gIleGlnCysCysGluThrValAsnValCysArgAspIleThrArgProProLys----- 2058

Qy    1239 CAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACA- 1297
                       |||   |||   |||  |||      :::        |||||
Db    2059 -ThrValAlaThrThrArgProThrProHis-ProThrGlyAlaGlnThrGlnThrThrP 2078

Qy    1298 -----------------AGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGG------- 1333
                           |||     |||      |||  :::|||  |||
Db    2078 heThrThrHisMetProSerAlaSerThrGluGlnProThrAlaThrSerArgGlyGlyP 2098

Qy    1334 ---------------------------------------------------TGCTGCC 1340
                                                            |||   |
Db    2098 roThrAlaThrSerValThrGlnGlyThrHisThrThrProValThrArgAsnCysHisP 2118

Qy    1341 CGCCATGCTCG---ATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCG------- 1390
           ||   |||:::        :::     |||     |||:::     |||
Db    2118 roArgCysThrTrpThrThrTrpPheAspValAspPheProSerProGlyProHisGlyG 2138

Qy    1391 ----------ATATATTCAAGAATGTAATCGA------CAAGAAATTCAAAGGCGACAC 1433
                        |||:::  :::    |||     ::::::|||  ||||||
Db    2138 lyAspLysGluThrTyrAsnAsnIle-IleArgSerGlyGluLysIleCysArgArgPro 2157

Qy    1434 GAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCA 1493
           ||||||::: ||||||:::
Db    2158 GluGluIleThrArgLeuGln--------------------------------------- 2164

Qy    1494 TGCCATGCTCAAGTCCATGGACAA------------------GGAAAAGTTTGCCAAGGC 1535
           |||  |||::: |||   :::                ||||||:::|||   |||
Db    2165 CysArgAlaLysSerHisProGluValSerIleGluHisLeuGlyGlnValValGln--- 2183

Qy    1536 TATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTAT 1595
                                                        |||||||||
Db    2184 ---------------------------------------------CysSerArgGlu 2187

Qy    1596 TCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGC 1655
            |||                :::|||       :::  |||
Db    2188 GluGly---------------LeuValCysArgAsnGlnAspGlnGlnGly--ProPheL 2202

Qy    1656 CGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCA------------- 1702
             :::|||:::             |||||||||        |||
Db    2202 ysMetCysLeuAsnIleGluValArgValLeuCysCysGluThrProLysGlyCysProV 2222

Qy    1703 --ACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGC 1760
             ||||||     |||     |||||||||||            |||        |
Db    2222 alThrSerThrProVal------ThrAlaProSerThrProSerGlyArgAlaIleSerP 2240

Qy    1761 CTGGTACAACTATCATACGACAGGCAAGGGCGTATTGG---AGAAGCAGGATCCTAAGAG 1817
           ||                 |||:::  :::   |||  :::||||||
Db    2240 ro---------------ThrGlnSerThrSerSerTrpGlnLysSerArgThrThrThrL 2255

Qy    1818 CGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTCG 1877
           ::::::                  :::||||||:::|||  |||:::|||
Db    2255 euValThr-----------------ThrSerThrThrSerThrProGlnThrSerThrT 2269

Qy    1878 CTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGG 1937
                      |||  |||:::  |||              |||||||||
Db    2269 hrTyrAlaHisThrThrSerThrThrSerAlaProThrAlaArgThrThrSer------- 2286

Qy    1938 TAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGG 1997
                      ||||||  :::     |||:::|||  :::        |||
Db    2287 -----------AlaProThrThrSerThrThrSerValProThrThr------SerThrI 2301

Qy    1998 CAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAG 2057
             :::|||              |||  |||      |||     |||     ||||
Db    2301 leSerGlyPro------LysThrThrProSerProValProThrThrSerThrThrSerA 2319

Qy    2058 CGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCAT 2117
           ||   |||::::::||| |||:::      |||:::     |||   |||  ::::
Db    2319 laAlaThrThrSerThrIleSerAlaProThrThrSerThrThrSerValProGlyThrT 2339

Qy    2118 CCAAGAGC 2125
           ::  |||
Db    2339 hrProSer 2341
```

RESULT 13
Q29071
ID   Q29071   PRELIMINARY;   PRT;   528 AA.
AC   Q29071;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Gastric mucin (Fragment).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX   MEDLINE=95275264; PubMed=7755593;
RA   Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA   LaMont J.T.;
RT   "Isolation and characterization of cDNA clones encoding pig gastric
RT   mucin.";
RL   Biochem. J. 308:89-96(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX   MEDLINE=94102478; PubMed=7506218;
RA   Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,

```
RA   LaMont J.T.;
RT   "Pig gastric mucin: isolation and characterization of a cDNA clone
RT   with a novel tandem repeat.";
RL   Gastroenterology 106:200-200(1994).
DR   EMBL; U10281; AAC48526.1; -.
FT   NON_TER       1      1
FT   NON_TER     528    528
SQ   SEQUENCE   528 AA;  49907 MW;  0BF0F6879203B2EA CRC64;

Alignment Scores:
Pred. No.:              0.0716       Length:         528
Score:                  149.00       Matches:        154
Percent Similarity:     34.22%       Conservative:   91
Best Local Similarity:  21.51%       Mismatches:     243
Query Match:            3.90%        Indels:         228
DB:                     6            Gaps:           28

US-10-008-355-1 (1-2139) x Q29071 (1-528)

Qy       68 AAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGC 127
               :::|||:::   :::      |||   |||   |||
Db        5 GlnProSerSerSerSerSerSerProThrThrSer------------------------ 16

Qy      128 GTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTG 187
               :::::::::::::|||                     |||:::|||      :::
Db       17 ---ThrThrSerValGlnSer--------------------SerSerSerSerVal 28

Qy      188 CCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCC 247
            |||:::|||      :::||||||                    |||       :::
Db       29 ProIleProSerThrThrSerValGln--------------------ProSerSerSer 41

Qy      248 TGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATC 307
               |||   |||||||||:::|||   :::|||
Db       42 GlySerAlaProThrThrSerAlaThrSerVal-------------------------- 52

Qy      308 ACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGG 367

Db       52 ------------------------------------------------------------ 52

Qy      368 GTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGC 427
                                                     :::|||          |||
Db       53 -------------------------------------GlnThrSerSerSerSerSer 59

Qy      428 TCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAAC 487
               :::|||   |||
Db       60 ProProIleSerSerThr------------------------------------------ 65

Qy      488 TGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCA 547
                           :::      :::|||:::|||:::|||         |||:::|||
Db       66 --------------IleSerValGlnThrSerSerSerSer---------SerValPro 77

Qy      548 ACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTC 607
            ||||||:::|||:::|||                                     :::
Db       78 ThrThrSerThrThrSer---------------------------------------Val 84

Qy      608 CTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACA 667
            |||:::         ||||||   :::      ||||||         |||   |||
Db       85 GlnProSer---------SerSerSerSerAlaProThrThr---------ArgAlaThr 98

Qy      668 CGGGCGACTTCAGCGTATTCCGCGTGTATGCCG-------------------------- 700
                     |||:::   |||:::      |||
Db       99 SerValGlnSerSerSerSerSerAlaProIleSerSerThrThrSerValGlnPro 118

Qy      701 ---------------GTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCT 745
                     |||||||||||          :::         :::   :::
Db      119 SerSerSerGlySerValProThrThrSerAlaThrSerValGlnSerSerSerSerSer 138

Qy      746 ATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCA 805
               :::|||   ||||||                     :::   :::   ::::::
Db      139 SerAlaProThrThrSer-----------AlaThrSerValGlnProSerSerSerSer 154

Qy      806 TGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATC 865
               |||   :::|||      |||   :::      |||
Db      155 SerProProIleSerSerThrValSerValGlnProSer-------------------- 167

Qy      866 GTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGA 925
               |||   :::           |||      :::|||           :::|||
Db      168 ---SerSerSerSerAlaProThrThrSerAlaThrSerValGlnProSerSerSerSer 186

Qy      926 AGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGA 985
               |||         |||         ||||||         :::|||
Db      187 SerProPro---------IleSerSerThrValSerValGlnThrSerSerSerSerSer 203

Qy      986 GTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGA 1045
            |||   |||                        |||:::   :::   |||
Db      204 ValProThr-----------------------ThrSerThrThrSerValGlnProSer 215

Qy     1046 TAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGA 1105
                     |||||||||                      |||||||||::::::
Db      216 SerSerSerSerValProThr---------ThrSerAlaThrSerValArgSerSerSer 232

Qy     1106 GTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGG 1165
                     |||      :::            |||   :::
Db      233 SerSerSerThrProIleProSerThrThrSerValGlnProSerSerSerSerSerAla 252

Qy     1166 CCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTT 1225
            ||||||                       :::||||||   |||
Db      253 ProThr-----------------ThrSerAlaThrSerValGlnProSerSerSerSer 266

Qy     1226 TTGCACAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAAT 1285
                     :::|||:::                  :::   |||   ||||||:::
Db      267 SerThrProIleProSer-----------ThrThrSerValGlnProSerSerSerSer 282

Qy     1286 CGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCA 1345
               |||   |||   ||||||   :::   :::                  ||||||
Db      283 SerAlaProThr---ThrSerAlaThrSerValGlnProSerSerSerSerSerProPro 301

Qy     1346 TGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATG 1405
               |||         :::   :::|||:::      ||||||
Db      302 IleSer---------SerThrIleSerValGlnProSerSer----------------- 312

Qy     1406 TAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGA 1465
               |||:::   :::         :::      |||:::|||   |||   |||
Db      313 ---SerSerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGlySer 331

Qy     1466 GTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGT 1525
                              |||||||||   |||   ||||||              |||
Db      332 AlaProThrThrSerAlaThrSerValGlnProSerSerSer--------------Ser 346

Qy     1526 TTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTG 1585
               |||   :::|||   |||
Db      347 SerProProIleSerSerThrIle----------------------------------- 354

Qy     1586 CTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTC 1645
                                                      ::::::   :::|||
Db      355 -------------------------------------SerValGlnProSerSerSer 361

Qy     1646 TTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACT 1705
               |||   |||                |||      |||            :::   :::
Db      362 SerSerSerPro------ThrThrSerThrThrSerValGlnProSerSerSerGlySer 379

Qy     1706 TCACCATGCGTATGAGCTACGGCT------CCATCAAGGGATATGAACCGCAGGACGGTG 1759
            :::|||         ||||||:::         ||||||         :::         |||
Db      380 AlaProThrThrSerAlaThrSerValGlnProSerSerSerSerValProThrThr 399

Qy     1760 CCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCG 1819
               |||::::::                        ||||||               :::
Db      400 SerAlaThrSerVal------------------------ArgSerSerSerSerSer 411

Qy     1820 ATGAGTTTGCCGTACAGG---------AGAATATCCTCGACCTCTTCCGCACCAAAAACT 1870
               :::|||            :::   ||||||:::||||||   |||   |||
Db      412 ThrProIleProThrThrThrSerValGlnProSerSerSerSerValProThrThr 431

Qy     1871 ATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCA 1930
                        :::|||   ||||||   |||               :::|||||||||
Db      432 SerAlaThrSerValGlnThrSerSerSerSerSerThrProIleProSerThrThrSer 451
```

```
Qy   1931 -------------------------------CGGGCGGTAACTCCGGTAGCCCCGTAT 1957
                                         :::   |||   |||   |||
Db    452 ValGlnProSerSerSerSerSerAlaProThrThrSerAlaThrSerValGlnProSer 471

Qy   1958 TCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTG 2017
           |||      :::         |||      :::|||:::
Db    472 SerSerSerSerProProIleSerSerThrIleSerValGlnProSer------------ 487

Qy   2018 GTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTC 2077
              :::|||||||||:::|||      |||      |||      |||::::::
Db    488 ---SerSerSerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGly 506

Qy   2078 TCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGC 2125
           |||:::      |||:::      |||         ||||||   |||
Db    507 SerAlaProThrThrSerAlaThrSerValGlnProSerSerSerSer 522
```

RESULT 14
Q8WZX1

```
ID   Q8WZX1      PRELIMINARY;      PRT;  1037 AA.
AC   Q8WZX1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Related to DNA repair endonuclease rad2.
GN   B7N14.040.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL669986; CAD21041.1; -.
DR   InterPro; IPR000513; Exo_N_I.
DR   InterPro; IPR002950; Josephin.
DR   InterPro; IPR003903; UIM.
DR   InterPro; IPR001532; XPGC_Rad.
DR   Pfam; PF02809; UIM; 2.
DR   Pfam; PF00867; XPG_I; 1.
DR   Pfam; PF00752; XPG_N; 1.
DR   SMART; SM00484; XPGI; 1.
DR   SMART; SM00485; XPGN; 1.
KW   Endonuclease.
SQ   SEQUENCE   1037 AA;  114115 MW;  C3977877D9444239 CRC64;

Alignment Scores:
Pred. No.:               0.139        Length:        1037
Score:                   147.00       Matches:       152
Percent Similarity:      31.96%       Conservative:  80
Best Local Similarity:   20.94%       Mismatches:    225
Query Match:             3.85%        Indels:        269
DB:                      3            Gaps:          38

US-10-008-355-1 (1-2139) x Q8WZX1 (1-1037)

Qy     58 GGGGTAGCCAAAGCCGACAAAGGC---------ATGTGG---------CTCCTCAACGAA 99
           |||:::   :::||||||:::            |||            |||     |||
Db    231 GlyIleArgArgAlaAspGluAlaAlaLeuGluArgTrpLysGluArgLeuThrArgGlu 250

Qy    100 CTC------AATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCTCCCG--- 150
           |||      |||:::               |||   |||            :::|||
Db    251 LeuArgThrAsnGluSerGlyPhePheArgThrArgHisLysAlaLeuAspIleProGlu 270

Qy    151 ------------TTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTG 198
                      |||:::   |||      :::      |||   ::::::   :::
Db    271 GlyPheProSerLeuGluValLeuArgTyrTyrThrHisProValValSerLysSerGlu 290

Qy    199 GTTATC------------------TTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
             :::                  |||   |||                  |||
Db    291 ThrValGlnArgLeuLysArgGlnPheProGly---------LysGlnAspValAspVal 307

Qy    241 CAGGGCCTG---ATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGC 297
           :::||||||   :::|||                                    :::
Db    308 GluGlyLeuArgValPheVal---------------------------------ArgGlu 316

Qy    298 ACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCT------CGCACGATGGGTGAG 351
           |||   |||   :::|||   |||   |||   |||         |||   :::
Db    317 ThrPheAspTrpGluTyr---ArgThrGlyAlaValLysPheIleArgValLeuAla--- 334

Qy    352 GAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGAC 411
                       |||   |||   |||::::::|||
Db    335 ------------ProSerLeuLeuValGlnPheLeu------------------LeuGln 344

Qy    412 AAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAGCTCAG 471
           :::   ||||||                |||:::   :::   ||||||:::   :::
Db    345 ArgSerGluGlyAsp---------------AspAspThrAspAspLeuArgArgLysGlu 359

Qy    472 GAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTA 531
                                  ||||||   |||                     |||
Db    360 ------------------------LysGluGluAla---------------AlaPheVal 366

Qy    532 GAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTT 591
           :::         |||         :::|||            |||         ::::::
Db    367 LysAlaIleLysSerArgArgThrHisPheSerThr------AspAlaThrProGluLeu 384

Qy    592 CGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGG 651
           |||:::   :::   |||
Db    385 ArgIleSerTyrValPro------------------------------------------ 390

Qy    652 ATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAAC 711
                                          :::         |||         |||
Db    391 ------------------------------IleLysValValThrLeuAspLeuAspAla 400

Qy    712 CGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTA 771
              |||   |||         :::|||||||:::   :::      |||      |||
Db    401 GluProGluGlu------GluAspAsnGluGlyPheGlyArgValGlyLeuAlaLeuAsn 418

Qy    772 TCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACG 831
           |||      :::   |||:::|||                     ||||||||||||
Db    419 SerAspAspGluPheAspAlaGluAspGluThr---------------ProGlySerThr 433

Qy    832 GAT--------------------------------------------------------- 834

Db    434 GlnGlyGlySerSerAlaLysProPheAspProPheLysProAspLeuAlaTrpIlePro 453

Qy    835 CGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATC 894
                 :::         ||||||      :::|||
Db    454 GluThrValAlaLysLeuGlyValProLeuThrValGlu--------------------- 466

Qy    895 GAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACC 954
                                  ||||||:::|||            :::|||
Db    467 ------------------------IleTrpGluGlu---------LysGlnArgAlaLys 475

Qy    955 CGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGT 1014
              :::|||   |||:::|||      :::   |||         |||   :::   |||
Db    476 GluLeuLysAlaAlaAsnLysGlyThrArgLysAlaGlnThrThrLysGlnThrGlyGly 495

Qy   1015 ATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCA 1074
           |||   |||                                                |||
Db    496 MetProAlaGly--------------------------------------------Ala 500

Qy   1075 TTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCT 1134
                 :::::::::               ||||||      |||   |||      |||
Db    501 LeuGluLysTyrValLysValThrLysGlySerAlaThrAlaGlyThrValThrLysSer 520

Qy   1135 CTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAG 1194
            :::         :::      :::            |||               |||
Db    521 -ThrSerAspTyrLeuSerLeuAspSerSerProProArgSerSerHisIleAlaProPr 540
```

```
Qy   1195 ACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACA 1254
            |||||||||  ||||||        |||          ||||||
Db    540 oArgSerSerTyrValPro------PheSerSerGlnGlnLeuPro------------- 553

Qy   1255 AATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTG-----------ACGACAAGTAC 1302
                          |||:::||||||:::|||             |||||||||
Db    554 ---------------ProLeuSerSerAspArgPheGluThrAlaThrThrThrSerLe 568

Qy   1303 AAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGC 1362
            |||||||||:::   :::   |||    :::
Db    568 uLysThrThrThrSerLysGlyThrThrLysAlaSer---------------------- 580

Qy   1363 CGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTC 1422
                                                     |||   |||   ||
Db    581 --------------------------------------------SerAlaArgThrSe 585

Qy   1423 AAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGC 1482
            ||||  |||                                                ||
Db    585 rLysLysThr-----------------------------------------------ThrAl 590

Qy   1483 GACAAGTTCCATGCCATGCTCAAGTCCATGGACA---------AGGAAAAGTTTGCCAAG 1533
            ||||||     |||     :::||||||:::              ::::::      ::
Db    590 aThrSerArgProProSerGlyAsnProTrpSerLeuAlaGlySerGlnAlaSerThrLy 610

Qy   1534 GCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCT 1593
            :::::::::   |||
Db    610 sIleThrLysAsnIleAlaSer----------------------------------- 617

Qy   1594 ATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCT-- 1651
                                              |||       :::   |||:::|||
Db    618 ------------------------SerGlnProAlaAlaLysThrSerLeuPheAspHi 629

Qy   1652 -TTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACC 1710
              |||:::   :::      :::|||   |||      :::    :::|||   |||||
Db    629 sGluProIleLeuIleSerSerSerProSerValProGlyLysArgLeuProAsnSerPr 649

Qy   1711 ATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCT-------- 1762
            |                     .          :::   ||||||       |||
Db    649 oLeu----------------------GluAsnAsnArgLeuPheSerProAspProPh 661

Qy   1763 -GGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATC------CTAAG 1815
             |||::::::     |||   |||       |||:::   |||:::           ||
Db    661 eGlySerSerProProArgIleAlaPro------TrpLysGluArgLeuProAlaThrAr 679

Qy   1816 AGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGT 1875
            |        |||   |||   |||       |||       |||         :::    ||
Db    679 gThrThrSerAspProThrArgHisProSerGlnArgSerMetArgAspGluGluProVa 699

Qy   1876 CGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGA------ACAACGACATC 1929
            |                    |||:::::::::   |||   |||
Db    699 lArgTrpArgSerGlnSerSerAlaLeuAlaSerSerProArgLysGlyProAlaPheGl 719

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTCGATA------------------------- 1963
            :::       ||||||         ||||||
Db    719 yLysAspLysThrProThrLysGlnArgSerIleLeuAspPheGlyTyrProSerThrTh 739

Qy   1964 ------------------------------AGAACGGCCGTCTGATCGGTCTTGCTTTC 1992
                                         ||||||     . |||    :::|||
Db    739 rLysGluGlnAlaGlySerArgLeuPheGlyArgThrGln------SerAlaValLeuPr 757

Qy   1993 GATGGCAACTGGGAAG 2008
                  |||||||||
Db    757 oSerThrThrGlyLys 762
```

RESULT 15
Q99ZW2

```
ID   Q99ZW2      PRELIMINARY;      PRT;   1368 AA.
AC   Q99ZW2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical protein SPy1046.
GN   SPY1046.
OS   Streptococcus pyogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC   Streptococcaceae; Streptococcus.
OX   NCBI_TaxID=1314;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX   MEDLINE=21192684; PubMed=11296296;
RA   Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA   Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA   Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J.,
RA   Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT   "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR   EMBL; AE006549; AAK33936.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   1368 AA;  158440 MW;  07D04F0B5965762F CRC64;

Alignment Scores:
Pred. No.:               0.173        Length:         1368
Score:                   146.50       Matches:        167
Percent Similarity:      31.74%       Conservative:   105
Best Local Similarity:   19.49%       Mismatches:     228
Query Match:             3.84%        Indels:         357
DB:                      16           Gaps:           43

US-10-008-355-1 (1-2139) x Q99ZW2 (1-1368)

Qy    103 AATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTC 162
          |||:::||||||   |||    :::    ::::::        ||||||   |||
Db    609 AsnGluGluAsnGluAspIleLeuGluAspIleValLeuThrLeuThrLeuPheGlu--- 627

Qy    163 TACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACC 222
                    |||:::   . |||          :::      :::
Db    628 ---------AspArgGluMetIleGluGluArgLeuLysThrTyrAlaHis--------- 641

Qy    223 GGTATCACAGTGTCCGATCAGGGCCTG------ATCTTTACCAACCACCACTGCGGATAC 276
                     |||:::   :::      :::       :::     |||:::
Db    642 ------LeuPheAspAspLysValMetLysGlnLeuLysArgArgArgTyrThrGlyTrp 659

Qy    277 GGTGCTATC---------------------------------CAGAGCCAAAGCACGGTG 303
           |||   :::                                   ||||||    |||:::
Db    660 GlyArgLeuSerArgLysLeuIleAsnGlyIleArgAspLysGlnSerGlyLysThrIle 679

Qy    304 GATCACGACTATCTGCGC---GATGGTTTCGTTTCTCGC--------------------- 339
                |||:::|||:::   |||||||||   :::|||
Db    680 ---LeuAspPheLeuLysSerAspGlyPheAlaAsnArgAsnPheMetGlnLeuIleHis 698

Qy    340 -----------ACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCC------------ 375
                     |||   |||::::::          :::|||
Db    699 AspAspSerLeuThrPheLysGluAspIleGlnLysAlaGlnValSerGlyGlnGlyAsp 718

Qy    376 ----------------------------------------------GTGAAGTATCTGCGC 390
                                                          |||   :::
Db    719 SerLeuHisGluHisIleAlaAsnLeuAlaGlySerProAlaIleLysLysGlyIleLeu 738

Qy    391 AAGATCGTGAAGGTAACGGAC---------AAGGTAGAAGGACAG-------------- 426
           :::   |||||||||   |||         ||||||   |||:::
Db    739 GlnThrValLysValValAspGluLeuValLysValMetGlyArgHisLysProGluAsn 758

Qy    427 -----------------------------------CTCAAGGGTATCACTGACGAGATG 450
                                                ||||||      :::
Db    759 IleValIleGluMetAlaArgGluAsnGlnThrThrGlnLysGlyGlnLysAsnSerArg 778

Qy    451 GAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAA---------AAAGAA 501
          ||||||:::::::::   :::|||      :::||||||           ||||||
Db    779 GluArgMetLysArgIleGluGluGlyIleLysGluLeuGlySerGlnIleLeuLysGlu 798

Qy    502 AATGCAGACGAGAAC---CAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATAC 558
          :::      ||||||   ||||||                       ||||||
Db    799 HisProValGluAsnThrGlnLeu-----------------------GlnAsnGluLys 810

Qy    559 TTCCTCATCGTCTAC--------------GATGTATTCAAGGACGTTCGTATGGTATTT 603
```

```
                   :::   |||             |||::::::   |||       :::
Db     811 LeuTyrLeuTyrTyrLeuGlnAsnGlyArgAspMetTyrValAspGlnGluLeuAspIle 830

Qy     604 GCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGT 663

Db     830 ------------------------------------------------------------ 830

Qy     664 CACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAA 723
                                                        ||||||   ::::::
Db     831 ---------------------------------------------AsnArgLeuSerAsp 835

Qy     724 TACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGC 783
           |||   |||:::                                  :::     |||
Db     836 TyrAspValAspHis------------------------------IleValProGlnSer 845

Qy     784 TACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTC 843
           :::   ||||||                              |||   |||      :::
Db     846 PheLeuLysAspAsp---------------------------SerIleAspAsnLysVal 856

Qy     844 ACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGC 903
              :::      ::::::|||   ::::::::|||   |||      ||||||
Db     857 LeuThrArgSerAspLysAsnArgGlyLysSerAspAsnValProSerGluGluValVal 876

Qy     904 GGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAA 963
                 :::      |||:::::   ::::::|||         |||:::   |||
Db     877 LysLysMetLysAsnTyrTrpArgGlnLeuLeuAsnAlaLysLeuIleThrGlnArgLys 896

Qy     964 TATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGC 1023
           :::                        |||      |||         |||
Db     897 Phe---------------------AspAsnLeuThrLysAlaGluArgGly--------- 906

Qy    1024 GGTCTCGCTCGTCTTGACGTGATAGGT---------CGTAAGCGTGCCGAGGAAAGAGCA 1074
           ||||||:::   ||||||      |||         |||:::      |||   |||
Db     907 GlyLeuSerGluLeuAspLysAlaGlyPheIleLysArgGlnLeuValGluThrArgGln 926

Qy    1075 TTCGCAGACTGGATC------------------CGTAAGAACGGCAAG------------ 1104
                 :::                        |||   |||   |||
Db     927 IleThrLysHisValAlaGlnIleLeuAspSerArgMetAsnThrLysTyrAspGluAsn 946

Qy    1104 ------------------------------------------------------------ 1104

Db     947 AspLysLeuIleArgGluValLysValIleThrLeuLysSerLysLeuValSerAspPhe 966

Qy    1104 ------------------------------------------------------------ 1104

Db     967 ArgLysAspPheGlnPheTyrLysValArgGluIleAsnAsnTyrHisHisAlaHisAsp 986

Qy    1105 ---------AGTGCTGTCTATGGCGATGTATTG------------TCTTCTCTCGAAAAG 1143
                    :::||||||   |||      |||                  ||||||
Db     987 AlaTyrLeuAsnAlaValValGlyThrAlaLeuIleLysLysTyrProLysLeuGluSer 1006

Qy    1144 GCTTATAAGGAAGGA---------------------------------GCCAAGGCCAAC 1170
              :::   |||                                    ||||||:::
Db    1007 GluPheValTyrGlyAspTyrLysValTyrAspValArgLysMetIleAlaLysSerGlu 1026

Qy    1171 CGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCA 1230
           :::|||:::               :::      |||   :::   ::::::   |||
Db    1027 GlnGluIleGlyLysAlaThrAlaLysTyrPhePheTyrSerAsnIleMetAsnPhePhe 1046

Qy    1231 CAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCG--- 1287
           :::               ||||||                     ::::::::::::
Db    1047 LysThrGluIleThrLeuAlaAsnGlyGluIleArgLysArgProLeuIleGluThrAsn 1066

Qy    1288 ------------------CTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGT 1329
                             :::   ||||||   :::|||:::         :::   |||
Db    1067 GlyGluThrGlyGluIleValTrpAspLysGlyArgAspPhe------AlaThrValArg 1084

Qy    1330 AAGGTGCTGCCC---GCCATGCTCGATATTGTACGCCGGCGT------------------ 1368
           |||||||||               ::::::||||||::::::
Db    1085 LysValLeuSerMetProGlnValAsnIleValLysLysThrGluValGlnThrGlyGly 1104

Qy    1369 ------------------ATCCCT---------GCCGACAAGCTCCCCGATATATTCAAG 1401
                             :::|||         :::|||||||||
Db    1105 PheSerLysGluSerIleLeuProLysArgAsnSerAspLysLeu--------------- 1119

Qy    1402 AATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGAC 1461
                 |||   :::|||         |||   |||||||||         ||||||
Db    1120 ------IleAlaArgLysLysAspTrpAspProLysLysTyrGlyGly------PheAsp 1135

Qy    1462 AAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCC------------------------ 1497
                    |||   ||||||            |||
Db    1136 SerProThrValAlaTyrSerValLeuValValAlaLysValGluLysGlyLysSerLys 1155

Qy    1498 ATGCTCAAGTCC---------------------------ATGGACAAGGAAAAGTTTGCC 1530
              |||||||||                           |||::::::      |||
Db    1156 LysLeuLysSerValLysGluLeuLeuGlyIleThrIleMetGluArgSerSerPheGlu 1175

Qy    1531 ------------------------------------AAGGCTATCGAGAAAGATCCGGCA 1554
                                                |||   ::::::||||||
Db    1176 LysAsnProIleAspPheLeuGluAlaLysGlyTyrLysGluValLysLysAspLeuIle 1195

Qy    1555 GTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCC 1614
           ::::::|||   |||
Db    1196 IleLysLeuProLys------------------------------------------Tyr 1201

Qy    1615 AATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTAC 1674
           :::   :::   :::|||   |||::::::               |||      |||:::
Db    1202 SerLeuPheGluLeuGluAsnGlyArgLysArgMetLeuAlaSerAlaGlyGluLeuGln 1221

Qy    1675 CCCGGACGT------GCTCTGCCGAGCGAT---GCCAACTTCACCATGCGTATGAGC--- 1722
              |||         ||||||||||||         ||||||            |||
Db    1222 LysGlyAsnGluLeuAlaLeuProSerLysTyrValAsnPheLeuTyrLeuAlaSerHis 1241

Qy    1723 TACGGCTCCATCAAGGGATATGAACCGCAGGACGGT------------GCCTGGTACAAC 1770
           |||      :::||||||      |||:::|||                  :::
Db    1242 TyrGluLysLeuLysGly---SerProGluAspAsnGluGlnLysGlnLeuPheValGlu 1260

Qy    1771 TATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCC 1830
              |||         |||   |||::::::         :::      ||||||:::
Db    1261 GlnHis---------LysHisTyrLeuAspGluIleIleGluGlnIleSerGluPheSer 1277

Qy    1831 GTACAG------------------------------------------------------ 1836
              :::
Db    1278 LysArgValIleLeuAlaAspAlaAsnLeuAspLysValLeuSerAlaTyrAsnLysHis 1297

Qy    1837 ---------------------------GAGAATATCCTCGACCTCTTCCGCACCAAAAAC 1869
                                      |||||||||:::   ||||||         |||
Db    1298 ArgAspLysProIleArgGluGlnAlaGluAsnIleIleHisLeuPheThrLeuThrAsn 1317

Qy    1870 TATGGT---------------------------------------------CGCTATGCC 1884
              |||                                            ||||||
Db    1318 LeuGlyAlaProAlaAlaPheLysTyrPheAspThrThrIleAspArgLysArgTyrThr 1337

Qy    1885 GAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGC 1935
                    ::::::      |||   |||   :::      |||||||||
Db    1338 SerThrLysGluValLeuAspAlaThrLeuIleHisGlnSerIleThrGly 1354
```

Search completed: May 16, 2003, 12:35:34
Job time : 221.5 secs

D
D

GenCore version 5.1.4_p5_4578


```
OM nucleic - nucleic search, using sw model

Run on:         May 23, 2003, 04:53:13 ; Search time 3810 Seconds
                                          (without alignments)
                                          16338.819 Million cell updates/sec

Title:          US-10-008-355-1
Perfect score:  2139
Sequence:       1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:  OLIGO_NUC
                Gapop 60.0 , Gapext 60.0

Searched:       2054640 seqs, 14551402878 residues

Word size :      0

Total number of hits satisfying chosen parameters:      4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21	1.0	5898	8	AY038360	AY038360 Nectria h
	2	21	1.0	165743	2	AL691459	AL691459 Homo sapi
	3	21	1.0	166143	9	AC013447	AC013447 Homo sapi
	4	21	1.0	200368	9	AC096541	AC096541 Homo sapi
c	5	20	0.9	4309	2	AC017213	AC017213 Drosophil
c	6	20	0.9	10689	1	AE004008	AE004008 Xylella f
c	7	20	0.9	86519	9	AP001860	AP001860 Homo sapi
c	8	20	0.9	129240	9	AC083826	AC083826 Homo sapi
	9	20	0.9	131813	2	AC011067	AC011067 Drosophil
c	10	20	0.9	136241	2	AC120223	AC120223 Rattus no
c	11	20	0.9	150505	2	AC095129	AC095129 Rattus no
c	12	20	0.9	152120	9	AL391294	AL391294 Human DNA
c	13	20	0.9	160279	2	AC019024	AC019024 Homo sapi
	14	20	0.9	160282	3	AC009253	AC009253 Drosophil
	15	20	0.9	163706	2	AC098197	AC098197 Rattus no
c	16	20	0.9	165837	9	AC005552	AC005552 Homo sapi
c	17	20	0.9	167700	2	AC079735	AC079735 Mus muscu
c	18	20	0.9	168944	2	AC125392	AC125392 Pan trogl
	19	20	0.9	169891	2	AC104241	AC104241 Homo sapi
c	20	20	0.9	171225	2	AC128439	AC128439 Rattus no
c	21	20	0.9	172715	10	CNS07EGP	AL591826 BAC 13C18
c	22	20	0.9	177024	2	AC102780	AC102780 Mus muscu
c	23	20	0.9	177262	2	AC123979	AC123979 Papio cyn
c	24	20	0.9	183238	2	AC053525	AC053525 Homo sapi
c	25	20	0.9	183915	2	AC073967	AC073967 Homo sapi
	26	20	0.9	186640	2	AC129152	AC129152 Rattus no
	27	20	0.9	187681	2	AC080124	AC080124 Homo sapi
	28	20	0.9	193134	2	AC016702	AC016702 Homo sapi
	29	20	0.9	195538	9	AC092733	AC092733 Homo sapi
c	30	20	0.9	196423	2	AC116933	AC116933 Papio cyn
	31	20	0.9	210636	9	AC006443	AC006443 Homo sapi
	32	20	0.9	220848	2	AC091579	AC091579 Homo sapi
c	33	20	0.9	223078	9	AL512504	AL512504 Human DNA
c	34	20	0.9	228944	2	AL844530	AL844530 Mus muscu
	35	20	0.9	272605	3	AE003669	AE003669 Drosophil
	36	20	0.9	274875	2	AC127582	AC127582 Mus muscu
c	37	20	0.9	297876	2	AC124538	AC124538 Mus muscu
	38	19	0.9	148	5	AY044428	AY044428 Oncorhync
	39	19	0.9	218	5	AF104614	AF104614 Salmo sal
	40	19	0.9	237	5	AF104589	AF104589 Oncorhync
	41	19	0.9	237	5	AF104590	AF104590 Oncorhync
	42	19	0.9	237	5	AF104607	AF104607 Salmo sal
	43	19	0.9	240	5	AF104608	AF104608 Salmo sal
	44	19	0.9	240	5	AF104610	AF104610 Salmo sal
	45	19	0.9	240	5	OTU80294	U80294 Oncorhynchu

ALIGNMENTS

```
RESULT 1
AY038360
LOCUS       AY038360                5898 bp    DNA     linear   PLN 09-AUG-2002
DEFINITION  Nectria haematococca mpVI retrotransposon Nht2, complete sequence.
ACCESSION   AY038360
VERSION     AY038360.1  GI:14700031
KEYWORDS    .
SOURCE      Nectria haematococca mpVI.
  ORGANISM  Nectria haematococca mpVI
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreales; Nectriaceae; Nectria.
REFERENCE   1  (bases 1 to 5898)
  AUTHORS   Shiflett,A.M., Enkerli,J. and Covert,S.F.
  TITLE     Nht2, a copia LTR retrotransposon from a conditionally dispensable
            chromosome in Nectria haematococca
```

```
  JOURNAL   Curr. Genet. 41 (2), 99-106 (2002)
  MEDLINE   22067396
   PUBMED   12073091
REFERENCE   2  (bases 1 to 5898)
  AUTHORS   Shiflett,A.M., Enkerli,J. and Covert,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2001) Forest Resources, University of Georgia,
            Brooks Drive, Athens, GA 30602-2152, USA
FEATURES             Location/Qualifiers
     source          1..5898
                     /organism="Nectria haematococca mpVI"
                     /isolate="156-30-6"
                     /db_xref="taxon:70791"
                     /chromosome="MAK1"
                     /note="mpVI"
     repeat_region   1..5898
                     /note="degenerate copia-like retrotransposon"
                     /transposon="Nht2"
     repeat_region   1..5
                     /note="5' duplicated target site"
                     /rpt_type=direct
     LTR             6..233
     misc_feature    <3984..>5537
                     /note="similar to reverse trancriptase and RNase H
                     domains"
     LTR             5666..5893
     repeat_region   5894..5898
                     /note="3' duplicated target site"
                     /rpt_type=direct
BASE COUNT     1766 a   1483 c   1265 g   1383 t      1 others
ORIGIN

  Query Match             1.0%;  Score 21;  DB 8;  Length 5898;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1137 CGAAAAGGCTTATAAGGAAGG 1157
           |||||||||||||||||||||
Db    1864 CGAAAAGGCTTATAAGGAAGG 1884


RESULT 2
AL691459
LOCUS       AL691459               165743 bp    DNA     linear   HTG 17-AUG-2002
DEFINITION  Homo sapiens chromosome 1 clone RP11-543D5, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL691459
VERSION     AL691459.24  GI:22415898
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165743)
  AUTHORS   Hopkins,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 21, 2002 this sequence version replaced gi:22265419.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA543D5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator; 15% of reads
            Chemistry: Dye-terminator Big Dye; 84% of reads
            Consensus quality: 165711 bases at least Q40
            Consensus quality: 165743 bases at least Q30
            Consensus quality: 165743 bases at least Q20
            Insert size: 165743; sum-of-contigs
            Insert size: 179024; 7.4% error; agarose-fp
            Quality coverage: 11.08x in Q20 bases; sum-of-contigs Quality
            coverage: 10.63x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..165743
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-543D5"
                     /clone_lib="RPCI-11.2"
     misc_feature    1..165743
                     /note="assembly_fragment:02071
                     clone_end:T7
                     vector_side:left"
BASE COUNT    41605 a  39965 c  40507 g  43666 t
ORIGIN

  Query Match             1.0%;  Score 21;  DB 2;  Length 165743;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     338 GCACGATGGGTGAGGAGCTTC 358
           |||||||||||||||||||||
Db  134330 GCACGATGGGTGAGGAGCTTC 134350


RESULT 3
AC013447
LOCUS       AC013447               166143 bp    DNA     linear   PRI 09-MAY-2001
DEFINITION  Homo sapiens BAC clone RP11-543D5 from 1, complete sequence.
ACCESSION   AC013447
VERSION     AC013447.4  GI:10334968
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166143)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 166143)
  AUTHORS   Joshu,C., Stoneking,T., Gregory,S. and Burkett,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-543D5
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 166143)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 166143)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 166143)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 166143)
```

```
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 28, 2000 this sequence version replaced gi:7631057.
            ------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            ------------- Summary Statistics
                Center project name: H_NH0543D05
            -------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-689L22.  Actual start of
            this clone is at base position 1 of RP11-543D5; actual end is at
            base position 166143 of RP11-543D5.

            The sequence fidelity of RP11-543D5 from base position 100749 to
            101626 can not be guaranteed.  This region is made up of a variable
            GT run where there are single M13 and PCR only regions.
FEATURES             Location/Qualifiers
     source          1..166143
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="1"
                     /clone="RP11-543D5"
                     /clone_lib="RPCI-11"
     repeat_region   79..574
                     /rpt_family="L2"
     repeat_region   2943..3362
                     /rpt_family="L2"
     repeat_region   3930..4010
                     /rpt_family="MIR"
     repeat_region   4107..4194
                     /rpt_family="MIR"
     repeat_region   4410..4766
                     /rpt_family="L2"
     repeat_region   4783..4857
                     /rpt_family="MIR"
     repeat_region   5456..5581
                     /rpt_family="L2"
     repeat_region   5631..5737
                     /rpt_family="CR1"
     repeat_region   6897..6961
                     /rpt_family="L1"
     repeat_region   7348..7465
                     /rpt_family="MIR"
     repeat_region   7567..7703
                     /rpt_family="MIR"
     repeat_region   9706..9809
                     /rpt_family="CR1"
     repeat_region   10546..10648
                     /rpt_family="L2"
     repeat_region   10649..10945
                     /rpt_family="Alu"
     repeat_region   10946..11080
                     /rpt_family="L2"
     repeat_region   11081..11462
                     /rpt_family="MaLR"
     repeat_region   11463..11526
                     /rpt_family="L2"
     repeat_region   11774..11894
                     /rpt_family="MIR"
     repeat_region   12407..12616
                     /rpt_family="MIR"
     repeat_region   13849..13950
                     /rpt_family="L2"
     repeat_region   15240..15467
                     /rpt_family="MIR"
     repeat_region   15520..15634
                     /rpt_family="MIR"
     repeat_region   15812..15969
                     /rpt_family="L2"
     repeat_region   16021..16139
                     /rpt_family="L2"
     repeat_region   16457..16677
                     /rpt_family="MIR"
     repeat_region   17300..17592
                     /rpt_family="L1"
     repeat_region   17826..17960
                     /rpt_family="L1"
     repeat_region   18236..18379
                     /rpt_family="MIR"
     repeat_region   18565..18980
                     /rpt_family="ERVL"
     repeat_region   18993..19107
                     /rpt_family="MIR"
     repeat_region   19181..19473
                     /rpt_family="Alu"
     repeat_region   19532..19756
                     /rpt_family="L1"
     repeat_region   19757..19930
                     /rpt_family="Alu"
     repeat_region   19931..19996
                     /rpt_family="L1"
     repeat_region   20075..20213
                     /rpt_family="MIR"
     repeat_region   20328..20518
                     /rpt_family="L2"
     repeat_region   20638..20929
                     /rpt_family="MER1_type"
     repeat_region   20958..21229
                     /rpt_family="MaLR"
     repeat_region   21230..21277
                     /rpt_family="L2"
     repeat_region   21449..21587
                     /rpt_family="MaLR"
     repeat_region   21618..21668
                     /rpt_family="L2"
     repeat_region   22345..22655
                     /rpt_family="Alu"
     repeat_region   22821..23158
                     /rpt_family="L2"
     repeat_region   23309..23379
```

```
                         /rpt_family="MIR"
     repeat_region       23705..23875
                         /rpt_family="MaLR"
     repeat_region       24042..24152
                         /rpt_family="MaLR"
     repeat_region       24440..24732
                         /rpt_family="Alu"
     repeat_region       24841..25234
                         /rpt_family="MaLR"
     repeat_region       25384..25642
                         /rpt_family="MIR"
     repeat_region       25694..25846
                         /rpt_family="MIR"
     repeat_region       26149..26345
                         /rpt_family="MIR"
     repeat_region       26440..26539
                         /rpt_family="L2"

  Query Match             1.0%;  Score 21;  DB 9;  Length 166143;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        338 GCACGATGGGTGAGGAGCTTC 358
              |||||||||||||||||||||
Db     134730 GCACGATGGGTGAGGAGCTTC 134750


RESULT 4
AC096541
LOCUS       AC096541              200368 bp    DNA     linear   PRI 31-MAY-2002
DEFINITION  Homo sapiens chromosome 1 clone RP11-330M19, complete sequence.
ACCESSION   AC096541 AL359084
VERSION     AC096541.2  GI:21281550
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200368)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200368)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 200368)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On May 31, 2002 this sequence version replaced gi:15638689.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: SC
            -------------- Project Information
                Center project name: chr-1
                Center clone name: RP11-330M19 (sc0667)
            ------------- Summary Statistics
                Sequencing vector: plasmid; 38% of reads
                Sequencing vector: plasmid; L08752; 62% of reads
                Chemistry: Dye-terminator ET; 89% of reads
                Chemistry: Dye-terminator Big Dye; 11% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 199860 bases at least Q40
                Consensus quality: 200007 bases at least Q30
                Consensus quality: 200368 bases at least Q20
                Insert size: 200368; sum-of-contigs
                Quality coverage: 9.5x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
                5':  Mapping in progress
                3':  RP11-307O17 AL451079
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
                  HindIII                EcoRI                 BglII
            -------------------  -------------------  -------------------
            SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt
            ---------  --------  ---------  --------  ---------  --------

            ---------            ---------            ---------
                  859       862       8696      8787       2761      2748
            ---------            ---------            ---------
                 6382      6512          6      <800       2067      2051
            ---------            ---------            ---------
                  512      <800        169      <800       9556      9558
            ---------            ---------            ---------
                  449      <800       7794      7749        751       797
            ---------            ---------            ---------
                 5425      5304       1389      1365       1632      1611
            ---------            ---------            ---------
                 1893      1888       1144      1129       5843      5875
            ---------            ---------            ---------
                  538      <800       8353      8399       5875      5875
            ---------            ---------            ---------
                  567      <800       1209      1199        100      <800
            ---------            ---------            ---------
                 4932      4924       4077      4005       6408      6406
            ---------            ---------            ---------
                 3378      3413       1150      1129       7217      7267
            ---------            ---------            ---------
                 1519      1502        505      <800        108      <800
            ---------            ---------            ---------
                 1788      1776       3804      3782       4723      4657
            ---------            ---------            ---------
                 4138      4136       8058      8027        255      <800
            ---------            ---------            ---------
                  468      <800        975       964        708      <800
            ---------            ---------            ---------
                  709      <800        470      <800        549      <800
```

2417	2375	859	885	4278	4222
1016	1010	681	<800	840	797
4965	4924	9699	9654	16573	16486
1391	1362	1738	1747	4181	4222
1828	1776	5846	5852	344	<800
474	<800	7651	7749	9127	9122
5588	5553	7	<800	1212	1203
1009	1010	3687	3782	4184	4222
8222	8258	1676	1747	794	797
4161	4136	649	<800	2549	2589
2143	2146	4518	4510	4070	4033
13940	13696	991	1014	20135	20286
7650	7663	13108	12755	3615	3595
994	1010	59	<800	1653	1611
2664	2660	3792	3782	3390	3401
67	<800	1237	1199	4250	4222
4553	4490	905	964	925	934
1617	1598	139	<800	5490	5551
727	722	6393	6300	197	<800
12486	12224	2138	2106	3602	3595
2982	3036	1987	1970	218	<800
5411	5304	823	830	9719	9558
3361	3413	1978	1970	588	<800
7890	7901	2403	2398	7908	8047
2581	2660	2790	2833	1766	1766
5622	5553	13902	13932	11220	10965
128	<800	2548	2512	61	<800
232	<800	61	<800	960	934
84	<800	2996	3009	6639	6761
13376	13696	1777	1747	2476	2589
8244	8258	1039	1014	88	<800
1510	1502	12719	12755	4639	4657
5987	6004	389	<800	3630	3595
1609	1598	339	<800	10219	10176
2327	2375	5609	5615	3324	3401
13583	13696	1789	1747	296	<800

3040	3036	5425	5378	908	934
6507	6512	4189	4224	2599	2589
2751	2807	6624	6730	252	<800
2303	2375	2208	2224	1460	1441
344	<800	11408	11348	135	<800

```
  Query Match             1.0%;  Score 21;  DB 9;  Length 200368;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      338 GCACGATGGGTGAGGAGCTTC 358
            |||||||||||||||||||||
Db    12725 GCACGATGGGTGAGGAGCTTC 12745


RESULT 5
AC017213/c
LOCUS       AC017213                4309 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017213
VERSION     AC017213.1  GI:6553773
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 4309)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10209899 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..4309
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     1270 a    846 c    867 g   1326 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 2;  Length 4309;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 GCAAATGAAATTAAAAAGTA 22
            ||||||||||||||||||||
Db     2113 GCAAATGAAATTAAAAAGTA 2094


RESULT 6
AE004008/c
LOCUS       AE004008               10689 bp    DNA     linear   BCT 15-JUN-2001
DEFINITION  Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.
ACCESSION   AE004008 AE003849
VERSION     AE004008.1  GI:9106961
KEYWORDS    .
SOURCE      Xylella fastidiosa 9a5c.
  ORGANISM  Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 10689)
  AUTHORS   Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
```

```
            Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
   PUBMED   10910347
REFERENCE   2  (bases 1 to 10689)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
            Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
            13083-970, Brazil
FEATURES             Location/Qualifiers
     source          1. .10689
                     /organism="Xylella fastidiosa 9a5c"
                     /db_xref="taxon:160492"
                     /clone="9a5c"
     gene            complement(119. .367)
                     /gene="XF1871"
     CDS             complement(119. .367)
                     /gene="XF1871"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84677.1"
                     /db_xref="GI:9106962"
                     /translation="MTMTLKEDKTEIIDEQRIQIAKLLSETAKIQAEIHEVNAHTQKL
                     MKETLKLTTESKWYPVVVGSSLIAAGATAATLFIKLFH"
     gene            complement(430. .618)
                     /gene="XF1872"
     CDS             complement(430. .618)
                     /gene="XF1872"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84678.1"
                     /db_xref="GI:9106963"
                     /translation="MMEEARATAPQGGLAASDPQRITDAPKKTGLRKRKYPTTGSASI
                     IVLRHVLTQRQQKRKASP"
     gene            941. .1288
                     /gene="XF1873"
     CDS             941. .1288
                     /gene="XF1873"
                     /note="similar to GI|479340 (percent identity: 36 %/query
                     alignment coverage:  87.8 %/subject alignment coverage:
                     90.2 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF84679.1"
                     /db_xref="GI:9106964"
                     /translation="MSKNKIPNQILTGNWAGFQFINGKLVTPERRTIEEWQLRWLSLT
                     CTLAREWQKMMEEARATAPQGGLAASDPQRITDALQKTVSHKRRHAANETASIIQLRQ
                     VLIQRQQKRKASP"
     gene            1351. .1662
                     /gene="XF1874"
     CDS             1351. .1662
                     /gene="XF1874"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84680.1"
                     /db_xref="GI:9106965"
                     /translation="MTMTLKEDKTEIIDEQKIQIAKLISETAKIQAEIHEVTTHTQEM
                     NAHINEMNARTQEMNARTQKLIEETQKVSKESQWYPVVVASGLLAAGATAATLFIKLF
                     H"
     gene            complement(1839. .2918)
                     /gene="XF1875"
     CDS             complement(1839. .2918)
                     /gene="XF1875"
                     /note="similar to SP|P03626 (percent identity: 36 %/query
                     alignment coverage:  88.0 %/subject alignment coverage:
                     105.0 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="phage-related protein"
                     /protein_id="AAF84681.1"
                     /db_xref="GI:9106966"
                     /translation="MLNLVTGAPGNGKTLYAVDWLIRQIEIDKSLVKSGAVPRSFYTD
                     IEGFDVEAVRRLTGYVVQSAPEDWRTTPQGSVIVYDEAHRMFPAGRPGRSDDPRVCDL
                     DTHRHGGYDLMFVTQWPTKIHHELRRLVGEHVHLNRAMGLQTAGLYRWSRAQDDPYDI
                     HQREKAEEEVWKFPKDRYALYASSTLHTVSHKFRIPKKVWSALSVCVTCSVIGLVFWH
                     YYSPAHLSEAASSVAGASGQASLRAAPASLSSSRSLVSGMRTYAVLETESAPTLSGCV
                     SSDRSCRCFNTDGYQIDMSVVECRRLLASPLPFNVYHAYVTSSSSAAAVSSSPSSSPS
                     SSSVSLSSSVSSAPILGSSSVPSSH"
     gene            complement(2918. .3202)
                     /gene="XF1876"
     CDS             complement(2918. .3202)
                     /gene="XF1876"
                     /note="similar to SP|P03625 (percent identity: 36 %/query
                     alignment coverage:  85.1 %/subject alignment coverage:
                     86.0 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="phage-related protein"
                     /protein_id="AAF84682.1"
                     /db_xref="GI:9106967"
                     /translation="MFSWLAMLLRNILGQTVARGLVGAGLALVTTVPLIPLVTSALNL
                     IVSKMSGISADVLNIALLMGFGEALSIIGSAMLTRLALQSLHVGIVKATT"
     gene            complement(3195. .4412)
                     /gene="XF1877"
     CDS             complement(3195. .4412)
```

```
                     /gene="XF1877"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84683.1"
                     /db_xref="GI:9106968"
                     /translation="MCSRSFFLYFISALGLTLVISFPFCLFAQNVPPTVEVVPTSVRY
                     SSVISDGVSTSATFEARSTALVNGVRYYTVPVDISASTLGSLAKVAVRRGMAFYNVYS
                     MLSGLINGAGWVIDELTHEVMSGPALKEIPVGTVAWWWQRPGDGHVFYSVTPQGLIAP
                     INAYQSSLSPPQPLVTSSGPSSVSSERWLYHLEGGGEYIGHLSKTDQSVPDYSSGLPP
                     SVILDTDLGQLVRSDPSTVNAVLTDSQTGAVLLTPEIVSALNKLRRSLEDELKASHAP
                     DQQPSSGGASSPPSSSGTAWPSFCSWASVVCDFIDWVKSDEFLKKPLVPPDVPYVDKL
                     PQAKTWSSGLGEGACPSPTTIPIEFSGYKTSVEISYQPFCDFAALMRPVVIVIATILA
                     AYIAGGFRGVKNV"
     gene            complement(4481..4702)
                     /gene="XF1878"
     CDS             complement(4481..4702)
                     /gene="XF1878"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84684.1"
                     /db_xref="GI:9106969"
                     /translation="MKSYIDRGIAFFSLLSTSSLVFAADAAGASFDAGAAVTALGAIS
                     GAVALIGAAKLAPAAISVGWKWLKGAIFG"
     gene            complement(4724..4924)
                     /gene="XF1879"
     CDS             complement(4724..4924)
                     /gene="XF1879"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84685.1"
                     /db_xref="GI:9106970"
                     /translation="MAQQLMSLYCTQYDVEARTCSQQAWMVPPSLLPPISYEDVRILL
                     PHIVMCFLVAWGFHFLFTVVRD"
     gene            complement(4931..5236)

  Query Match             0.9%;  Score 20;  DB 1;  Length 10689;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        646 AACTGGATGTGGCCGCGTCA 665
              ||||||||||||||||||||
Db       9863 AACTGGATGTGGCCGCGTCA 9844


RESULT 7
AP001860/c
LOCUS       AP001860               86519 bp    DNA     linear   PRI 20-MAR-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 4q22-q24, clone:343A4,
            complete sequence.
ACCESSION   AP001860
VERSION     AP001860.2  GI:13111481
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:343A4.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Tsai,S.F.
  TITLE     Genome Sequencing of the Chromosome 4q Region Implicated in Human
            Hepatocellular Carcinoma Pathogenesis 1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 86519)
  AUTHORS   Tsai,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2000) Shih-Feng Tsai, National Yang-Ming
            University, Institute of Genetics; 155 Li-Rong St. Section 2,
            Peitou, Taipei, Taiwan 11221, Republic of China
            (E-mail:ympetsai@ym.edu.tw, URL:http://genome.ym.edu.tw/,
            Tel:886-2-28267043, Fax:886-2-28264930)
COMMENT     On Feb 22, 2001 this sequence version replaced gi:7649705.
            Quality: the expected Phred/Phrap calculated error rate (per 10 kb)
            is 0.0019; Estimated total number of errors is 0.0167.
FEATURES             Location/Qualifiers
     source          1..86519
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4q22-q24"
                     /clone="343A4"
BASE COUNT    29441 a  16281 c  16491 g  24306 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 9;  Length 86519;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          8 TGAAATTAAAAAGTATTCTT 27
              ||||||||||||||||||||
Db      36681 TGAAATTAAAAAGTATTCTT 36662


RESULT 8
AC083826/c
LOCUS       AC083826              129240 bp    DNA     linear   PRI 07-NOV-2001
DEFINITION  Homo sapiens BAC clone RP11-289C17 from 4, complete sequence.
ACCESSION   AC083826
VERSION     AC083826.7  GI:14589704
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129240)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 129240)
  AUTHORS   VanBrunt,A. and Meyer,R.
  TITLE     The sequence of Homo sapiens BAC clone RP11-289C17
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 129240)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 129240)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 129240)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jul 3, 2001 this sequence version replaced gi:14269820.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0289C17
```

```
-------------.

NOTICE:  This sequence may not represent the entire insert of this
clone.  It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO.  For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries.  Genomics 51:1-8.  The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR:  pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-10L12, 2000 bp overlap; the
clone sequenced to the right is AC022793.  Actual start of this
clone is at base position 156151 of RP11-10L12; actual end is at
base position 129240 of RP11-289C17.

Data from AC08124 was used to finish this clone, AC083826.  The
region from 45337 to 45361 is only covered by data from AC08124.
Polymorphisms exist between RP11-289C17 and RP11-10L12.
FEATURES             Location/Qualifiers
     source          1..129240
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-289C17"
                     /clone_lib="RPCI-11"
     repeat_region   25..309
                     /rpt_family="Alu"
     misc_feature    222..662
                     /note="match to EST AI791382 (NID:g5339098) oh01a11.y5"
     misc_feature    273..756
                     /note="match to EST AI732988 (NID:g5054101) oh01a11.x5"
     repeat_region   949..1263
                     /rpt_family="Alu"
     repeat_region   1234..1274
                     /rpt_family="(A)n"
     repeat_region   1310..1593
                     /rpt_family="L1"
     repeat_region   1594..1870
                     /rpt_family="Alu"
     repeat_region   1871..2198
                     /rpt_family="L1"
     misc_feature    2223..2226
                     /note="similar to Homo sapiens EST AV650986
                     (NID:g9872000)"
     misc_feature    2223..2226
                     /note="similar to Homo sapiens EST BG291978
                     (NID:g13050347)"
     repeat_region   2471..2626
                     /rpt_family="L1"
     repeat_region   2627..2900
                     /rpt_family="Alu"
     repeat_region   2921..2948
                     /rpt_family="AT_rich"
     repeat_region   3597..3662
                     /rpt_family="L2"
     repeat_region   3814..4312
                     /rpt_family="ERV"
     repeat_region   4361..4495
                     /rpt_family="L1"
     repeat_region   4501..4684
                     /rpt_family="ERV1"
     misc_feature    4644
                     /note="similar to Homo sapiens EST BG386439
                     (NID:g13279885)"
     misc_feature    4681..5216
                     /note="similar to Mus musculus EST AA059627 (NID:g1553413)
                     mj74g02.r1"
     misc_feature    4681..5087
                     /note="similar to Mus musculus EST BF462055
                     (NID:g11531238)"
     misc_feature    4681..5040
                     /note="similar to Homo sapiens EST BF906415
                     (NID:g12297873)"
     misc_feature    4681..4992
                     /note="similar to Bos taurus EST BE484374 (NID:g9603907)"
     misc_feature    4697..5115
                     /note="similar to Homo sapiens EST BG386439
                     (NID:g13279885)"
     misc_feature    4701..5382
                     /note="similar to Homo sapiens EST BG576912
                     (NID:g13584565)"
     misc_feature    4724
                     /note="similar to Homo sapiens EST BG827066
                     (NID:g14174653)"
     misc_feature    4728..5272
                     /note="similar to Homo sapiens EST BG178667
                     (NID:g12685370)"
     misc_feature    4728..5000
                     /note="similar to Homo sapiens EST BE296600
                     (NID:g9180164)"
     misc_feature    4734..5379
                     /note="similar to Homo sapiens EST AU133548
                     (NID:g10994087)"
     misc_feature    4734..5275
                     /note="similar to Homo sapiens EST BG757135
                     (NID:g14067788)"
     misc_feature    4734..5179
                     /note="similar to Homo sapiens EST BE018798 (NID:g8278841)
                     bb85b04.y1"
     misc_feature    4734..5095
                     /note="similar to Homo sapiens EST BF527247
                     (NID:g11614610)"
     misc_feature    4743..5542
                     /note="similar to Mus musculus EST BG923425
                     (NID:g14303901)"
     misc_feature    4858..5275
                     /note="similar to Homo sapiens EST BG828340
                     (NID:g14175927)"
     misc_feature    4859..5379
                     /note="similar to Homo sapiens EST BG740954
                     (NID:g14051607)"
     misc_feature    4859..5357
                     /note="similar to Homo sapiens EST BG290663
                     (NID:g13047851)"
     misc_feature    4859..5115
                     /note="similar to Homo sapiens EST AW367838
                     (NID:g6872488)"
     misc_feature    4860..5383
                     /note="similar to Homo sapiens EST AW860531
                     (NID:g7956224)"
     misc_feature    4860..5115
                     /note="similar to Homo sapiens EST BE007807
                     (NID:g8268040)"
```

```
    misc_feature    4865. .5343
                    /note="similar to Homo sapiens EST BE779572
                    (NID:g10200770)"
    misc_feature    4870. .5387
                    /note="similar to Homo sapiens EST BG283401
                    (NID:g13033252)"
    misc_feature    4874. .5379
                    /note="similar to Homo sapiens EST AW991862
                    (NID:g8251932)"
    misc_feature    4875. .5383

  Query Match              0.9%;  Score 20;  DB 9;  Length 129240;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 TGAAATTAAAAAGTATTCTT 27
             ||||||||||||||||||||
Db   12472 TGAAATTAAAAAGTATTCTT 12453


RESULT 9
AC011067
LOCUS       AC011067              131813 bp    DNA     linear   HTG 20-MAR-2000
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR09L20 (D1140)
            RPCI-98 09.L.20 map 39A-39A strain y; cn bw sp, *** SEQUENCING IN
            PROGRESS ***, 130 unordered pieces.
ACCESSION   AC011067
VERSION     AC011067.4  GI:7264763
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 131813)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 131813)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Mar 20, 2000 this sequence version replaced gi:6984326.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 130 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      803: contig of 803 bp in length
```

```
            *      804      883: gap of unknown length
            *      884     1597: contig of 714 bp in length
            *     1598     1677: gap of unknown length
            *     1678     2236: contig of 559 bp in length
            *     2237     2316: gap of unknown length
            *     2317     2921: contig of 605 bp in length
            *     2922     3001: gap of unknown length
            *     3002     3804: contig of 803 bp in length
            *     3805     3884: gap of unknown length
            *     3885     4416: contig of 532 bp in length
            *     4417     4496: gap of unknown length
            *     4497     5056: contig of 560 bp in length
            *     5057     5136: gap of unknown length
            *     5137     5947: contig of 811 bp in length
            *     5948     6027: gap of unknown length
            *     6028     6767: contig of 740 bp in length
            *     6768     6847: gap of unknown length
            *     6848     7486: contig of 639 bp in length
            *     7487     7566: gap of unknown length
            *     7567     8210: contig of 644 bp in length
            *     8211     8290: gap of unknown length
            *     8291     9033: contig of 743 bp in length
            *     9034     9113: gap of unknown length
            *     9114     9770: contig of 657 bp in length
            *     9771     9850: gap of unknown length
            *     9851    10597: contig of 747 bp in length
            *    10598    10677: gap of unknown length
            *    10678    11128: contig of 451 bp in length
            *    11129    11208: gap of unknown length
            *    11209    12081: contig of 873 bp in length
            *    12082    12161: gap of unknown length
            *    12162    13148: contig of 987 bp in length
            *    13149    13228: gap of unknown length
            *    13229    14305: contig of 1077 bp in length
            *    14306    14385: gap of unknown length
            *    14386    15384: contig of 999 bp in length
            *    15385    15464: gap of unknown length
            *    15465    16434: contig of 970 bp in length
            *    16435    16514: gap of unknown length
            *    16515    17933: contig of 1419 bp in length
            *    17934    18013: gap of unknown length
            *    18014    18888: contig of 875 bp in length
            *    18889    18968: gap of unknown length
            *    18969    19673: contig of 705 bp in length
            *    19674    19753: gap of unknown length
            *    19754    20536: contig of 783 bp in length
            *    20537    20616: gap of unknown length
            *    20617    21346: contig of 730 bp in length
            *    21347    21426: gap of unknown length
            *    21427    22580: contig of 1154 bp in length
            *    22581    22660: gap of unknown length
            *    22661    23567: contig of 907 bp in length
            *    23568    23647: gap of unknown length
            *    23648    24529: contig of 882 bp in length
            *    24530    24609: gap of unknown length
            *    24610    25686: contig of 1077 bp in length
            *    25687    25766: gap of unknown length
            *    25767    27466: contig of 1700 bp in length
            *    27467    27546: gap of unknown length
            *    27547    29292: contig of 1746 bp in length
            *    29293    29372: gap of unknown length
            *    29373    30275: contig of 903 bp in length
            *    30276    30355: gap of unknown length
            *    30356    31407: contig of 1052 bp in length
            *    31408    31487: gap of unknown length
            *    31488    32729: contig of 1242 bp in length
            *    32730    32809: gap of unknown length
            *    32810    34227: contig of 1418 bp in length
            *    34228    34307: gap of unknown length
            *    34308    35359: contig of 1052 bp in length
            *    35360    35439: gap of unknown length
            *    35440    37213: contig of 1774 bp in length
            *    37214    37293: gap of unknown length
```

```
*     37294     38442: contig of 1149 bp in length
*     38443     38522: gap of unknown length
*     38523     39967: contig of 1445 bp in length
*     39968     40047: gap of unknown length
*     40048     41551: contig of 1504 bp in length
*     41552     41631: gap of unknown length
*     41632     43684: contig of 2053 bp in length
*     43685     43764: gap of unknown length
*     43765     45714: contig of 1950 bp in length
*     45715     45794: gap of unknown length
*     45795     47443: contig of 1649 bp in length
*     47444     47523: gap of unknown length
*     47524     49494: contig of 1971 bp in length
*     49495     49574: gap of unknown length
*     49575     52237: contig of 2663 bp in length
*     52238     52317: gap of unknown length
*     52318     53865: contig of 1548 bp in length
*     53866     53945: gap of unknown length
*     53946     55202: contig of 1257 bp in length
*     55203     55282: gap of unknown length
*     55283     57891: contig of 2609 bp in length
*     57892     57971: gap of unknown length
*     57972     59715: contig of 1744 bp in length
*     59716     59795: gap of unknown length
*     59796     63092: contig of 3297 bp in length
*     63093     63172: gap of unknown length
*     63173     66602: contig of 3430 bp in length
*     66603     66682: gap of unknown length
*     66683     70903: contig of 4221 bp in length
*     70904     70983: gap of unknown length
*     70984     74368: contig of 3385 bp in length
*     74369     74448: gap of unknown length
*     74449     80276: contig of 5828 bp in length
*     80277     80356: gap of unknown length
*     80357     80704: contig of 348 bp in length
*     80705     80784: gap of unknown length
*     80785     81593: contig of 809 bp in length
*     81594     81673: gap of unknown length
*     81674     82414: contig of 741 bp in length
*     82415     82494: gap of unknown length
*     82495     82954: contig of 460 bp in length
*     82955     83034: gap of unknown length
*     83035     83456: contig of 422 bp in length
*     83457     83536: gap of unknown length
*     83537     84100: contig of 564 bp in length
*     84101     84180: gap of unknown length
*     84181     84685: contig of 505 bp in length
*     84686     84765: gap of unknown length
*     84766     85313: contig of 548 bp in length
*     85314     85393: gap of unknown length
*     85394     86028: contig of 635 bp in length
*     86029     86108: gap of unknown length
*     86109     86727: contig of 619 bp in length
*     86728     86807: gap of unknown length
*     86808     87354: contig of 547 bp in length
*     87355     87434: gap of unknown length
*     87435     88012: contig of 578 bp in length
*     88013     88092: gap of unknown length
*     88093     88532: contig of 440 bp in length
*     88533     88612: gap of unknown length
*     88613     89345: contig of 733 bp in length
*     89346     89425: gap of unknown length
*     89426     90007: contig of 582 bp in length
*     90008     90087: gap of unknown length
*     90088     90949: contig of 862 bp in length
*     90950     91029: gap of unknown length
*     91030     91496: contig of 467 bp in length
*     91497     91576: gap of unknown length
*     91577     92074: contig of 498 bp in length
*     92075     92154: gap of unknown length
*     92155     92748: contig of 594 bp in length
*     92749     92828: gap of unknown length
*     92829     93411: contig of 583 bp in length
*     93412     93491: gap of unknown length
*     93492     94063: contig of 572 bp in length
*     94064     94143: gap of unknown length

  Query Match             0.9%;  Score 20;  DB 2;  Length 131813;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 GCAAATGAAATTAAAAAGTA 22
              ||||||||||||||||||||
Db      10083 GCAAATGAAATTAAAAAGTA 10102

RESULT 10
AC120223/c
LOCUS       AC120223              136241 bp    DNA     linear   HTG 18-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-307M22, *** SEQUENCING IN PROGRESS
            ***, 50 unordered pieces.
ACCESSION   AC120223
VERSION     AC120223.2  GI:21746475
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 136241)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 136241)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2002) Human Genome Sequencing Center, Department
```

```
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 136241)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 14, 2002 this sequence version replaced gi:20452765.
            ------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
                Center project name: GUML
                Center clone name: CH230-307M22
            ------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 92457 bases at least Q40
                Consensus quality: 97846 bases at least Q30
                Consensus quality: 101277 bases at least Q20
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 50 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1237: contig of 1237 bp in length
            *     1238      1337: gap of unknown length
            *     1338      2477: contig of 1140 bp in length
            *     2478      2577: gap of unknown length
            *     2578      3773: contig of 1196 bp in length
            *     3774      3873: gap of unknown length
            *     3874      5333: contig of 1460 bp in length
            *     5334      5433: gap of unknown length
            *     5434      6516: contig of 1083 bp in length
            *     6517      6616: gap of unknown length
            *     6617      7797: contig of 1181 bp in length
            *     7798      7897: gap of unknown length
            *     7898      9520: contig of 1623 bp in length
            *     9521      9620: gap of unknown length
            *     9621     11451: contig of 1831 bp in length
            *    11452     11551: gap of unknown length
            *    11552     12773: contig of 1222 bp in length
            *    12774     12873: gap of unknown length
            *    12874     14506: contig of 1633 bp in length
            *    14507     14606: gap of unknown length
            *    14607     15953: contig of 1347 bp in length
            *    15954     16053: gap of unknown length
            *    16054     18349: contig of 2296 bp in length
            *    18350     18449: gap of unknown length
            *    18450     19976: contig of 1527 bp in length
            *    19977     20076: gap of unknown length
            *    20077     21148: contig of 1072 bp in length
            *    21149     21248: gap of unknown length
            *    21249     22929: contig of 1681 bp in length
            *    22930     23029: gap of unknown length
            *    23030     25641: contig of 2612 bp in length
            *    25642     25741: gap of unknown length
            *    25742     28054: contig of 2313 bp in length
            *    28055     28154: gap of unknown length
            *    28155     30379: contig of 2225 bp in length
            *    30380     30479: gap of unknown length
            *    30480     32748: contig of 2269 bp in length
            *    32749     32848: gap of unknown length
            *    32849     35106: contig of 2258 bp in length
            *    35107     35206: gap of unknown length
            *    35207     37048: contig of 1842 bp in length
            *    37049     37148: gap of unknown length
            *    37149     39266: contig of 2118 bp in length
            *    39267     39366: gap of unknown length
            *    39367     40799: contig of 1433 bp in length
            *    40800     40899: gap of unknown length
            *    40900     43019: contig of 2120 bp in length
            *    43020     43119: gap of unknown length
            *    43120     45158: contig of 2039 bp in length
            *    45159     45258: gap of unknown length
            *    45259     47295: contig of 2037 bp in length
            *    47296     47395: gap of unknown length
            *    47396     48834: contig of 1439 bp in length
            *    48835     48934: gap of unknown length
            *    48935     51435: contig of 2501 bp in length
            *    51436     51535: gap of unknown length
            *    51536     54235: contig of 2700 bp in length
            *    54236     54335: gap of unknown length
            *    54336     57329: contig of 2994 bp in length
            *    57330     57429: gap of unknown length
            *    57430     59496: contig of 2067 bp in length
            *    59497     59596: gap of unknown length
            *    59597     61948: contig of 2352 bp in length
            *    61949     62048: gap of unknown length
            *    62049     63956: contig of 1908 bp in length
            *    63957     64056: gap of unknown length
            *    64057     66236: contig of 2180 bp in length
            *    66237     66336: gap of unknown length
            *    66337     69688: contig of 3352 bp in length
            *    69689     69788: gap of unknown length
            *    69789     72181: contig of 2393 bp in length
            *    72182     72281: gap of unknown length
            *    72282     75455: contig of 3174 bp in length
            *    75456     75555: gap of unknown length
            *    75556     78895: contig of 3340 bp in length
            *    78896     78995: gap of unknown length
            *    78996     82341: contig of 3346 bp in length
            *    82342     82441: gap of unknown length
            *    82442     84796: contig of 2355 bp in length
            *    84797     84896: gap of unknown length
            *    84897     88837: contig of 3941 bp in length
            *    88838     88937: gap of unknown length
            *    88938     93012: contig of 4075 bp in length
            *    93013     93112: gap of unknown length
            *    93113     97714: contig of 4602 bp in length
            *    97715     97814: gap of unknown length
            *    97815    100110: contig of 2296 bp in length
            *   100111    100210: gap of unknown length
            *   100211    105379: contig of 5169 bp in length
            *   105380    105479: gap of unknown length
            *   105480    111300: contig of 5821 bp in length
            *   111301    111400: gap of unknown length
            *   111401    116430: contig of 5030 bp in length
            *   116431    116530: gap of unknown length
            *   116531    121578: contig of 5048 bp in length
            *   121579    121678: gap of unknown length
            *   121679    128057: contig of 6379 bp in length
            *   128058    128157: gap of unknown length
            *   128158    136241: contig of 8084 bp in length.
FEATURES             Location/Qualifiers
     source          1..136241
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-307M22"
BASE COUNT    39106 a  27003 c  27430 g  37671 t   5031 others
ORIGIN

  Query Match             0.9%;  Score 20;  DB 2;  Length 136241;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       5 AAATGAAATTAAAAAGTATT 24
           ||||||||||||||||||||
Db  120229 AAATGAAATTAAAAAGTATT 120210


RESULT 11
AC095129/c
LOCUS       AC095129              150505 bp    DNA     linear   HTG 10-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-8E13, *** SEQUENCING IN PROGRESS ***,
            73 unordered pieces.
ACCESSION   AC095129
VERSION     AC095129.3  GI:21716997
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 150505)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150505)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 150505)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 9, 2002 this sequence version replaced gi:17942055.
            ------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
                Center project name: GCQH
                Center clone name: CH230-8E13
            ------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 83548 bases at least Q40
                Consensus quality: 89611 bases at least Q30
                Consensus quality: 94789 bases at least Q20
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 73 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1       1132: contig of 1132 bp in length
            *       1133       1232: gap of unknown length
            *       1233       2418: contig of 1186 bp in length
            *       2419       2518: gap of unknown length
            *       2519       3844: contig of 1326 bp in length
            *       3845       3944: gap of unknown length
            *       3945       5112: contig of 1168 bp in length
            *       5113       5212: gap of unknown length
            *       5213       6372: contig of 1160 bp in length
            *       6373       6472: gap of unknown length
            *       6473       7527: contig of 1055 bp in length
            *       7528       7627: gap of unknown length
            *       7628       8951: contig of 1324 bp in length
            *       8952       9051: gap of unknown length
            *       9052      10064: contig of 1013 bp in length
            *      10065      10164: gap of unknown length
            *      10165      11599: contig of 1435 bp in length
            *      11600      11699: gap of unknown length
            *      11700      13070: contig of 1371 bp in length
            *      13071      13170: gap of unknown length
            *      13171      14296: contig of 1126 bp in length
            *      14297      14396: gap of unknown length
            *      14397      15437: contig of 1041 bp in length
            *      15438      15537: gap of unknown length
            *      15538      16813: contig of 1276 bp in length
            *      16814      16913: gap of unknown length
            *      16914      17995: contig of 1082 bp in length
            *      17996      18095: gap of unknown length
            *      18096      19561: contig of 1466 bp in length
            *      19562      19661: gap of unknown length
            *      19662      20855: contig of 1194 bp in length
            *      20856      20955: gap of unknown length
            *      20956      22182: contig of 1227 bp in length
            *      22183      22282: gap of unknown length
            *      22283      23293: contig of 1011 bp in length
            *      23294      23393: gap of unknown length
            *      23394      24562: contig of 1169 bp in length
            *      24563      24662: gap of unknown length
            *      24663      25677: contig of 1015 bp in length
            *      25678      25777: gap of unknown length
            *      25778      27011: contig of 1234 bp in length
            *      27012      27111: gap of unknown length
            *      27112      28391: contig of 1280 bp in length
            *      28392      28491: gap of unknown length
            *      28492      29785: contig of 1294 bp in length
            *      29786      29885: gap of unknown length
```

```
*    29886    31450: contig of 1565 bp in length
*    31451    31550: gap of unknown length
*    31551    33272: contig of 1722 bp in length
*    33273    33372: gap of unknown length
*    33373    34554: contig of 1182 bp in length
*    34555    34654: gap of unknown length
*    34655    35908: contig of 1254 bp in length
*    35909    36008: gap of unknown length
*    36009    37552: contig of 1544 bp in length
*    37553    37652: gap of unknown length
*    37653    39518: contig of 1866 bp in length
*    39519    39618: gap of unknown length
*    39619    41678: contig of 2060 bp in length
*    41679    41778: gap of unknown length
*    41779    43376: contig of 1598 bp in length
*    43377    43476: gap of unknown length
*    43477    45110: contig of 1634 bp in length
*    45111    45210: gap of unknown length
*    45211    46627: contig of 1417 bp in length
*    46628    46727: gap of unknown length
*    46728    48564: contig of 1837 bp in length
*    48565    48664: gap of unknown length
*    48665    49903: contig of 1239 bp in length
*    49904    50003: gap of unknown length
*    50004    51077: contig of 1074 bp in length
*    51078    51177: gap of unknown length
*    51178    52965: contig of 1788 bp in length
*    52966    53065: gap of unknown length
*    53066    55285: contig of 2220 bp in length
*    55286    55385: gap of unknown length
*    55386    56530: contig of 1145 bp in length
*    56531    56630: gap of unknown length
*    56631    57808: contig of 1178 bp in length
*    57809    57908: gap of unknown length
*    57909    59912: contig of 2004 bp in length
*    59913    60012: gap of unknown length
*    60013    61628: contig of 1616 bp in length
*    61629    61728: gap of unknown length
*    61729    63702: contig of 1974 bp in length
*    63703    63802: gap of unknown length
*    63803    65518: contig of 1716 bp in length
*    65519    65618: gap of unknown length
*    65619    67487: contig of 1869 bp in length
*    67488    67587: gap of unknown length
*    67588    69222: contig of 1635 bp in length
*    69223    69322: gap of unknown length
*    69323    71614: contig of 2292 bp in length
*    71615    71714: gap of unknown length
*    71715    74337: contig of 2623 bp in length
*    74338    74437: gap of unknown length
*    74438    76038: contig of 1601 bp in length
*    76039    76138: gap of unknown length
*    76139    78343: contig of 2205 bp in length
*    78344    78443: gap of unknown length
*    78444    80973: contig of 2530 bp in length
*    80974    81073: gap of unknown length
*    81074    83111: contig of 2038 bp in length
*    83112    83211: gap of unknown length
*    83212    85690: contig of 2479 bp in length
*    85691    85790: gap of unknown length
*    85791    88004: contig of 2214 bp in length

 Query Match             0.9%;  Score 20;  DB 2;  Length 150505;
 Best Local Similarity   100.0%;  Pred. No. 45;
 Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     504 TGCAGACGAGAACCAACTCT 523
           ||||||||||||||||||||
Db   31654 TGCAGACGAGAACCAACTCT 31635

RESULT 12
AL391294/c
LOCUS       AL391294              152120 bp    DNA     linear   PRI 06-DEC-2001
DEFINITION  Human DNA sequence from clone RP11-196G4 on chromosome X Contains
            part of a novel gene and a CpG island, complete sequence.
ACCESSION   AL391294
VERSION     AL391294.16  GI:11121074
KEYWORDS    HTG; CpG island.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152120)
  AUTHORS   Clark,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 8, 2000 this sequence version replaced gi:11064146.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            This sequence is the entire insert of clone RP11-196G4 The true
            right end of clone RP13-104H2 is at 71449 in this sequence. This
            sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP11-196G4 is from
            the library RPCI-11.1 constructed by the group of Pieter de Jong.
            For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .152120
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-196G4"
                     /clone_lib="RPCI-11.1"
     repeat_region   1. .1563
                     /note="TIGGER1 repeat: matches 1. .1558 of consensus"
     repeat_region   1666. .1815
                     /note="75 copies 2 mer tt 58% conserved"
     repeat_region   2485. .2668
                     /note="MER5A repeat: matches 1. .184 of consensus"
     repeat_region   3055. .3100
                     /note="23 copies 2 mer tt 87% conserved"
     repeat_region   3809. .4297
                     /note="MER9 repeat: matches 1. .511 of consensus"
     repeat_region   5346. .5399
                     /note="27 copies 2 mer ac 98% conserved"
     repeat_region   5507. .6137
                     /note="L1M1 repeat: matches -1389. .-744 of consensus"
     repeat_region   6110. .6247
                     /note="L1M1 repeat: matches -1186. .-1037 of consensus"
     repeat_region   6231. .8974
                     /note="L1M1 repeat: matches -523. .2714 of consensus"
     misc_feature    complement(8868. .9026)
                     /note="match: GSS: Em:AZ392746"
     misc_feature    8952. .9312
```

```
                     /note="match: GSS: Em:AQ236903"
     misc_feature    complement(8971..9176)
                     /note="match: GSS: Em:AQ552982 Em:AQ568449"
     misc_feature    8971..9159
                     /note="match: STS: Em:Af002096"
     misc_feature    complement(9021..9179)
                     /note="match: STS: Em:G09176"
     repeat_region   9028..9179
                     /note="AluJb repeat: matches 138..289 of consensus"
     repeat_region   9180..12619
                     /note="L1MA2 repeat: matches 2325..5786 of consensus"
     repeat_region   12614..14953
                     /note="L1PA13 repeat: matches 3768..6152 of consensus"
     repeat_region   14961..15487
                     /note="L1MA2 repeat: matches 5769..6303 of consensus"
     gene            complement(15532..17999)
                     /gene="bA196G4.1"
     mRNA            complement(<15532..17999)
                     /gene="bA196G4.1"
                     /product="bA196G4.1 (novel protein)"
                     /note="match: cDNAs: Em:AK024111 Em:AK001535 Em:AL050030
                     match: ESTs: Em:AI252657 Em:BE673395 Em:AA947819
                     Em:BF142469 Em:AW883384 Em:BE005185 Em:BE180722
                     Em:BF141391 Em:AW892817 Em:BE926061 Em:BF139073 Em:F12610
                     Em:BE707336 Em:AA968512 Em:AA634946 Em:F08032 Em:BE781553
                     Em:BE018659 Em:W56397 Em:BE174649 Em:AI075935 Em:BE992345
                     Em:BF144949 Em:BE916748 Em:H17605 Em:AI694583 Em:AW524670
                     Em:AI798375 Em:BE005114 Em:AI110805 Em:BF140831
                     Em:AW369780 Em:AW885772 Em:AW885736 Em:BF056380
                     Em:AI065144 Em:R19847 Em:AW889619 Em:BE613917 Em:AW900587
                     Em:BE178770 Em:BE171589 Em:BE881716 Em:AA243338
                     Em:BE770557 Em:AA603412 Em:AA280341 Em:AW293265
                     Em:AW181892 Em:AW885540 Em:AW517792 Em:AA262381 Em:T06923
                     Em:AW993831 Em:BE163571 Em:BE170628 Em:BE709135
                     Em:BE564899 Em:BE175078 Em:F12118 Em:BE465085 Em:AI250884
                     Em:AA837657 Em:AA010153 Em:AA987568 Em:AW771325
                     Em:AW780434 Em:AW762521 Em:T25165 Em:AA242913 Em:W31490
                     Em:AW892807 Em:AF075304 Em:BE270146 Em:R57609 Em:AW879991
                     Em:BE915168 Em:AA598509 Em:BE827473 Em:AI572844
                     Em:BE827475 Em:AI370669 Em:AW900704 Em:AW270335
                     Em:BE569882 Em:AW771253 Em:AW660653"
                     /evidence=not_experimental
     repeat_region   15801..16081
                     /note="AluSc repeat: matches 1..281 of consensus"
     misc_feature    complement(16083..16344)
                     /gene="bA196G4.1"
                     /note="match: GSS: Em:AZ271635"
     misc_feature    complement(17589..17996)
                     /gene="bA196G4.1"
                     /note="match: GSS: Em:AQ710263"
     misc_feature    complement(18204..18563)
                     /note="match: GSS: Em:AQ730707"
     misc_feature    19122..19525
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   21236..21293
                     /note="MIR repeat: matches 85..142 of consensus"
     repeat_region   21550..21678
                     /note="MLT1A2 repeat: matches 220..349 of consensus"
     repeat_region   21681..24760
                     /note="L1PB2 repeat: matches 3013..6152 of consensus"
     misc_feature    21804..21855
                     /note="Sequence from overlapping clone AL359775. Assembly
                     confirmed by restriction digest"
     repeat_region   24754..25642
                     /note="L1P repeat: matches 1560..2455 of consensus"
     repeat_region   25634..26219
                     /note="L1P3 repeat: matches 8..416 of consensus"
     repeat_region   26225..26480
                     /note="L1MD1 repeat: matches 5566..5825 of consensus"
     repeat_region   26518..27573
                     /note="L1MEc repeat: matches 358..1420 of consensus"
     repeat_region   27597..27650
                     /note="THE1C-internal repeat: matches 1524..1576 of
                     consensus"
     repeat_region   27655..28025
                     /note="THE1C repeat: matches 1..371 of consensus"
     repeat_region   28026..28120
                     /note="FLAM_A repeat: matches 29..123 of consensus"
     repeat_region   28186..28209
                     /note="12 copies 2 mer ac 100% conserved"
     repeat_region   28331..28775
                     /note="MLT2A repeat: matches 9..453 of consensus"
     repeat_region   29122..29426
                     /note="AluSx repeat: matches 5..305 of consensus"
     repeat_region   29427..30036
                     /note="L1PA8 repeat: matches 10..602 of consensus"
     repeat_region   30037..30502
                     /note="L1PA8 repeat: matches 5696..6161 of consensus"
     repeat_region   30572..31103
                     /note="L1MD2 repeat: matches 5575..6116 of consensus"
     repeat_region   31104..31624
                     /note="L1M4 repeat: matches -2..556 of consensus"
     repeat_region   31629..34385
                     /note="L1P repeat: matches 1818..4658 of consensus"
     repeat_region   34390..34461
                     /note="24 copies 3 mer tat 68% conserved"
     repeat_region   34392..34463
                     /note="12 copies 6 mer ttatta 68% conserved"
     repeat_region   34500..34936
                     /note="L1PA8 repeat: matches 5724..6163 of consensus"
     repeat_region   35012..37203
                     /note="L1PA12 repeat: matches -355..1833 of consensus"
     repeat_region   37383..37682
                     /note="L1M2 repeat: matches -614..-321 of consensus"
     repeat_region   37755..38071
                     /note="L1M4 repeat: matches 4330..4650 of consensus"
     repeat_region   38052..38277
                     /note="L1PA12 repeat: matches 1323..1554 of consensus"
     repeat_region   38407..39569
                     /note="L1P repeat: matches 2047..3228 of consensus"
     repeat_region   39571..41509
                     /note="L1PA15 repeat: matches 4193..6148 of consensus"
     repeat_region   41502..42355
                     /note="L1 repeat: matches 3785..4644 of consensus"
     repeat_region   42366..48504
                     /note="L1PA4 repeat: matches 1..6141 of consensus"
     repeat_region   49131..49187
                     /note="MER93 repeat: matches 17..82 of consensus"
     repeat_region   49302..49911
                     /note="L1PA5 repeat: matches 5529..6138 of consensus"
     repeat_region   49921..50335
                     /note="L1PA5 repeat: matches 5135..5552 of consensus"
     repeat_region   50392..51778
                     /note="L1PA4 repeat: matches 2176..3577 of consensus"
     repeat_region   51779..52334
                     /note="LTR1 repeat: matches 194..785 of consensus"

  Query Match             0.9%;  Score 20;  DB 9;  Length 152120;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        485 AACTGGCCAAAAAAGAAAAT 504
              ||||||||||||||||||||
Db     139650 AACTGGCCAAAAAAGAAAAT 139631

RESULT 13
AC019024/c
LOCUS       AC019024               160279 bp    DNA     linear   HTG 14-MAR-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-478E10 map 11, WORKING DRAFT
            SEQUENCE, 3 ordered pieces.
ACCESSION   AC019024
VERSION     AC019024.5  GI:19424608
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens.
```

```
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1  (bases 1 to 160279)
  AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Homo sapiens chromosome 11, clone RP11-478E10
  JOURNAL Unpublished
REFERENCE 2  (bases 1 to 160279)
  AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
          Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
          Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
          Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
          DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
          Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
          Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
          Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
          Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
          Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
          McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
          Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
          Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
          Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
          Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
          Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (29-DEC-1999) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3  (bases 1 to 160279)
  AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
          Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
          Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
          Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
          Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
          Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
          Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
          Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
          MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
          McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
          Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
          Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
          Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
          Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
          Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
          Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
          Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (14-MAR-2002) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Mar 14, 2002 this sequence version replaced gi:19387766.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L5298
              Center clone name: 478_E_10
          -------------- Summary Statistics
              Sequencing vector: M13; M77815; 26% of reads
              Sequencing vector: Plasmid; n/a; 74% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 159553 bases at least Q40
              Consensus quality: 159829 bases at least Q30
              Consensus quality: 159947 bases at least Q20
              Insert size: 163000; agarose-fp
              Insert size: 160079; sum-of-contigs
              Quality coverage: 19.1 in Q20 bases; agarose-fp
              Quality coverage: 19.5 in Q20.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1  77856: contig of 77856 bp in length
          *    77857 77956: gap of       100 bp
          *    77957   79239: contig of 1283 bp in length
          *    79240 79339: gap of       100 bp
          *    79340   160279: contig of 80940 bp in length.
FEATURES             Location/Qualifiers
     source          1. .160279
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-478E10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .77856
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    77957. .79239
                     /note="assembly_fragment"
     misc_feature    79340. .160279
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    42570 a  40290 c  38390 g  38828 t    201 others
ORIGIN

  Query Match             0.9%;  Score 20;  DB 2;  Length 160279;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1135 CTCGAAAAGGCTTATAAGGA 1154
           ||||||||||||||||||||
Db   66837 CTCGAAAAGGCTTATAAGGA 66818


RESULT 14
AC009253
LOCUS       AC009253              160282 bp    DNA     linear   INV 08-NOV-2001
DEFINITION  Drosophila melanogaster, chromosome 2L, region 39A-39A, BAC clone
            BACR24F17, complete sequence.
ACCESSION   AC009253
VERSION     AC009253.16  GI:16798933
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 160282)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
```

```
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2L, region 39A-39A
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160282)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Nov 8, 2001 this sequence version replaced gi:7264755.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .160282
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="39A-39A"
                     /clone="BACR24F17 (D1038)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    47892 a   33440 c   31695 g   47255 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 3;  Length 160282;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 GCAAATGAAATTAAAAAGTA 22
              ||||||||||||||||||||
Db     3371 GCAAATGAAATTAAAAAGTA 3390


RESULT 15
AC098197
LOCUS       AC098197               163706 bp    DNA     linear   HTG 12-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-97I4, *** SEQUENCING IN PROGRESS ***,
            65 unordered pieces.
ACCESSION   AC098197
VERSION     AC098197.5  GI:21729815
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 163706)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163706)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 163706)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 11, 2002 this sequence version replaced gi:17973759.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GIKF
                Center clone name: CH230-97I4
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 116803 bases at least Q40
                Consensus quality: 121625 bases at least Q30
                Consensus quality: 125275 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 65 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
* be preserved.
*        1     1248: contig of 1248 bp in length
*     1249     1348: gap of unknown length
*     1349     2558: contig of 1210 bp in length
*     2559     2658: gap of unknown length
*     2659     3918: contig of 1260 bp in length
*     3919     4018: gap of unknown length
*     4019     5561: contig of 1543 bp in length
*     5562     5661: gap of unknown length
*     5662     7107: contig of 1446 bp in length
*     7108     7207: gap of unknown length
*     7208     8207: contig of 1000 bp in length
*     8208     8307: gap of unknown length
*     8308     9733: contig of 1426 bp in length
*     9734     9833: gap of unknown length
*     9834    10854: contig of 1021 bp in length
*    10855    10954: gap of unknown length
*    10955    12152: contig of 1198 bp in length
*    12153    12252: gap of unknown length
*    12253    13803: contig of 1551 bp in length
*    13804    13903: gap of unknown length
*    13904    14983: contig of 1080 bp in length
*    14984    15083: gap of unknown length
*    15084    16304: contig of 1221 bp in length
*    16305    16404: gap of unknown length
*    16405    17407: contig of 1003 bp in length
*    17408    17507: gap of unknown length
*    17508    18723: contig of 1216 bp in length
*    18724    18823: gap of unknown length
*    18824    19873: contig of 1050 bp in length
*    19874    19973: gap of unknown length
*    19974    21350: contig of 1377 bp in length
*    21351    21450: gap of unknown length
*    21451    22583: contig of 1133 bp in length
*    22584    22683: gap of unknown length
*    22684    23851: contig of 1168 bp in length
*    23852    23951: gap of unknown length
*    23952    25978: contig of 2027 bp in length
*    25979    26078: gap of unknown length
*    26079    27697: contig of 1619 bp in length
*    27698    27797: gap of unknown length
*    27798    28883: contig of 1086 bp in length
*    28884    28983: gap of unknown length
*    28984    30017: contig of 1034 bp in length
*    30018    30117: gap of unknown length
*    30118    31386: contig of 1269 bp in length
*    31387    31486: gap of unknown length
*    31487    33565: contig of 2079 bp in length
*    33566    33665: gap of unknown length
*    33666    35234: contig of 1569 bp in length
*    35235    35334: gap of unknown length
*    35335    36957: contig of 1623 bp in length
*    36958    37057: gap of unknown length
*    37058    38427: contig of 1370 bp in length
*    38428    38527: gap of unknown length
*    38528    39907: contig of 1380 bp in length
*    39908    40007: gap of unknown length
*    40008    41539: contig of 1532 bp in length
*    41540    41639: gap of unknown length
*    41640    43413: contig of 1774 bp in length
*    43414    43513: gap of unknown length
*    43514    45389: contig of 1876 bp in length
*    45390    45489: gap of unknown length
*    45490    47738: contig of 2249 bp in length
*    47739    47838: gap of unknown length
*    47839    49892: contig of 2054 bp in length
*    49893    49992: gap of unknown length
*    49993    52477: contig of 2485 bp in length
*    52478    52577: gap of unknown length
*    52578    54809: contig of 2232 bp in length
*    54810    54909: gap of unknown length
*    54910    56151: contig of 1242 bp in length
*    56152    56251: gap of unknown length
*    56252    58273: contig of 2022 bp in length
*    58274    58373: gap of unknown length
*    58374    61788: contig of 3415 bp in length
*    61789    61888: gap of unknown length
*    61889    64074: contig of 2186 bp in length
*    64075    64174: gap of unknown length
*    64175    67262: contig of 3088 bp in length
*    67263    67362: gap of unknown length
*    67363    70423: contig of 3061 bp in length
*    70424    70523: gap of unknown length
*    70524    74022: contig of 3499 bp in length
*    74023    74122: gap of unknown length
*    74123    76188: contig of 2066 bp in length
*    76189    76288: gap of unknown length
*    76289    78899: contig of 2611 bp in length
*    78900    78999: gap of unknown length
*    79000    80837: contig of 1838 bp in length
*    80838    80937: gap of unknown length
*    80938    83128: contig of 2191 bp in length
*    83129    83228: gap of unknown length
*    83229    85835: contig of 2607 bp in length
*    85836    85935: gap of unknown length
*    85936    88399: contig of 2464 bp in length
*    88400    88499: gap of unknown length
*    88500    92442: contig of 3943 bp in length
*    92443    92542: gap of unknown length
*    92543    96214: contig of 3672 bp in length
*    96215    96314: gap of unknown length
*    96315    99043: contig of 2729 bp in length
*    99044    99143: gap of unknown length
*    99144   103453: contig of 4310 bp in length
*   103454   103553: gap of unknown length
*   103554   106323: contig of 2770 bp in length
*   106324   106423: gap of unknown length
*   106424   110557: contig of 4134 bp in length

  Query Match             0.9%;  Score 20;  DB 2;  Length 163706;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1160 CCAAGGCCAACCGTGAGATG 1179
          ||||||||||||||||||||
Db 136347 CCAAGGCCAACCGTGAGATG 136366


Search completed: May 23, 2003, 08:42:14
Job time : 5892 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 23, 2003, 07:01:32 ; Search time 193 Seconds (without alignments) 14634.553 Million cell updates/sec

Title: US-10-008-355-1
Perfect score: 2139
Sequence: 1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 828747 seqs, 660231138 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2139	100.0	2139	9	US-10-008-355-1	Sequence 1, Appli
	2	19	0.9	490	9	US-09-918-995-37887	Sequence 37887, A
	3	19	0.9	505	9	US-09-918-995-37842	Sequence 37842, A
	4	19	0.9	4163	10	US-09-954-456-528	Sequence 528, App
	5	18	0.8	203	10	US-09-960-352-10103	Sequence 10103, A
c	6	18	0.8	252	10	US-09-974-300-7766	Sequence 7766, Ap
	7	18	0.8	355	10	US-09-960-352-12702	Sequence 12702, A
	8	18	0.8	408	10	US-09-960-352-9104	Sequence 9104, Ap
	9	18	0.8	410	10	US-09-960-352-10777	Sequence 10777, A
	10	18	0.8	417	10	US-09-960-352-5137	Sequence 5137, Ap
	11	18	0.8	418	10	US-09-960-352-11256	Sequence 11256, A
	12	18	0.8	427	10	US-09-960-352-12448	Sequence 12448, A
c	13	18	0.8	466	10	US-09-864-761-6022	Sequence 6022, Ap
	14	18	0.8	479	10	US-09-960-352-14538	Sequence 14538, A
	15	18	0.8	499	9	US-09-918-995-1375	Sequence 1375, Ap
	16	18	0.8	2699	8	US-08-834-666A-3	Sequence 3, Appli
	17	18	0.8	2798	8	US-08-834-666A-1	Sequence 1, Appli
	18	18	0.8	2915	8	US-08-834-666A-5	Sequence 5, Appli
	19	18	0.8	3995	10	US-09-919-497-9	Sequence 9, Appli
c	20	18	0.8	7557	12	US-10-044-090-253	Sequence 253, App
c	21	17	0.8	264	9	US-10-046-935-657	Sequence 657, App
c	22	17	0.8	264	9	US-09-878-178-657	Sequence 657, App
c	23	17	0.8	264	9	US-10-146-502-657	Sequence 657, App
	24	17	0.8	285	10	US-09-923-876-4571	Sequence 4571, Ap
	25	17	0.8	287	10	US-09-960-352-478	Sequence 478, App
c	26	17	0.8	348	9	US-09-918-995-18569	Sequence 18569, A
c	27	17	0.8	373	10	US-09-880-107-614	Sequence 614, App
c	28	17	0.8	399	10	US-09-960-253-126	Sequence 126, App
c	29	17	0.8	402	9	US-09-918-995-3910	Sequence 3910, Ap
	30	17	0.8	413	10	US-09-960-352-2414	Sequence 2414, Ap
	31	17	0.8	470	9	US-09-918-995-25073	Sequence 25073, A
	32	17	0.8	485	9	US-09-918-995-23928	Sequence 23928, A
	33	17	0.8	495	10	US-09-783-590-8450	Sequence 8450, Ap
c	34	17	0.8	516	10	US-09-919-580-852	Sequence 852, App
	35	17	0.8	600	10	US-09-917-800A-857	Sequence 857, App
	36	17	0.8	606	9	US-09-774-639-50	Sequence 50, Appl
	37	17	0.8	606	9	US-09-969-730-98	Sequence 98, Appl
c	38	17	0.8	623	9	US-09-883-152-29	Sequence 29, Appl
c	39	17	0.8	651	9	US-09-883-152-26	Sequence 26, Appl
c	40	17	0.8	687	9	US-09-774-639-106	Sequence 106, App
c	41	17	0.8	687	9	US-09-969-730-107	Sequence 107, App
	42	17	0.8	861	10	US-09-452-599-166	Sequence 166, App
	43	17	0.8	918	10	US-09-974-300-4593	Sequence 4593, Ap
	44	17	0.8	981	9	US-09-887-527-12	Sequence 12, Appl
c	45	17	0.8	998	9	US-10-001-857-87	Sequence 87, Appl

ALIGNMENTS

```
RESULT 1
US-10-008-355-1
; Sequence 1, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 2139
;   TYPE: DNA
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-1

  Query Match             100.0%;  Score 2139;  DB 9;  Length 2139;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2139;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60

Qy         61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120

Qy        121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180

Qy        181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
```

```
Qy     241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300

Qy     301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360

Qy     361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420

Qy     421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480

Qy     481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540

Qy     541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600

Qy     601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660

Qy     661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720

Qy     721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780

Qy     781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840

Qy     841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900

Qy     901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960

Qy     961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020

Qy    1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080

Qy    1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140

Qy    1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200

Qy    1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260

Qy    1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320

Qy    1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380

Qy    1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440

Qy    1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500

Qy    1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560

Qy    1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620

Qy    1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680

Qy    1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740

Qy    1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800

Qy    1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860

Qy    1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920

Qy    1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980

Qy    1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040

Qy    2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100

Qy    2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 2139
           |||||||||||||||||||||||||||||||||||||||
Db    2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 2139
```

```
RESULT 2
US-09-918-995-37887
; Sequence 37887, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 37887
;   LENGTH: 490
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(490)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-37887

  Query Match             0.9%;  Score 19;  DB 9;  Length 490;
  Best Local Similarity   100.0%;  Pred. No. 6.7;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      241 CAGGGCCTGATCTTTACCA 259
            |||||||||||||||||||
Db      274 CAGGGCCTGATCTTTACCA 292


RESULT 3
US-09-918-995-37842
; Sequence 37842, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 37842
;   LENGTH: 505
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(505)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-37842

  Query Match             0.9%;  Score 19;  DB 9;  Length 505;
  Best Local Similarity   100.0%;  Pred. No. 6.8;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      241 CAGGGCCTGATCTTTACCA 259
            |||||||||||||||||||
Db      310 CAGGGCCTGATCTTTACCA 328

RESULT 4
US-09-954-456-528
; Sequence 528, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 528
;   LENGTH: 4163
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-528

  Query Match             0.9%;  Score 19;  DB 10;  Length 4163;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      241 CAGGGCCTGATCTTTACCA 259
            |||||||||||||||||||
Db     3135 CAGGGCCTGATCTTTACCA 3153


RESULT 5
US-09-960-352-10103
; Sequence 10103, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10103
;   LENGTH: 203
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 43-LIB34-066-Q1-E1-C12
US-09-960-352-10103

  Query Match             0.8%;  Score 18;  DB 10;  Length 203;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1160 CCAAGGCCAACCGTGAGA 1177
            ||||||||||||||||||
Db       57 CCAAGGCCAACCGTGAGA 74


RESULT 6
US-09-974-300-7766/c
; Sequence 7766, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
```

```
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7766
;   LENGTH: 252
;   TYPE: DNA
;   ORGANISM: Bacillus clausii
US-09-974-300-7766

  Query Match             0.8%;  Score 18;  DB 10;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    639 TACGGACAACTGGATGTG 656
          ||||||||||||||||||
Db    159 TACGGACAACTGGATGTG 142

RESULT 7
US-09-960-352-12702
; Sequence 12702, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12702
;   LENGTH: 355
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 54-LIB34-010-Q1-E1-F10
US-09-960-352-12702

  Query Match             0.8%;  Score 18;  DB 10;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1160 CCAAGGCCAACCGTGAGA 1177
          ||||||||||||||||||
Db    335 CCAAGGCCAACCGTGAGA 352

RESULT 8
US-09-960-352-9104
; Sequence 9104, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 9104
;   LENGTH: 408
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 39-LIB3058-039-Q1-K1-B4
US-09-960-352-9104

  Query Match             0.8%;  Score 18;  DB 10;  Length 408;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1160 CCAAGGCCAACCGTGAGA 1177
          ||||||||||||||||||
Db    384 CCAAGGCCAACCGTGAGA 401

RESULT 9
US-09-960-352-10777
; Sequence 10777, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley.C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10777
;   LENGTH: 410
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (383)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: 46-LIB34-016-Q1-E1-D6
US-09-960-352-10777

  Query Match             0.8%;  Score 18;  DB 10;  Length 410;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1160 CCAAGGCCAACCGTGAGA 1177
          ||||||||||||||||||
Db    386 CCAAGGCCAACCGTGAGA 403

RESULT 10
US-09-960-352-5137
; Sequence 5137, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5137
;   LENGTH: 417
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 22-LIB34-030-Q1-E1-F5
US-09-960-352-5137

  Query Match             0.8%;  Score 18;  DB 10;  Length 417;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1160 CCAAGGCCAACCGTGAGA 1177
          ||||||||||||||||||
Db    117 CCAAGGCCAACCGTGAGA 134

RESULT 11
```

```
US-09-960-352-11256
; Sequence 11256, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11256
;   LENGTH: 418
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 48-LIB34-016-Q1-E1-D8
US-09-960-352-11256

  Query Match             0.8%;  Score 18;  DB 10;  Length 418;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1160 CCAAGGCCAACCGTGAGA 1177
          ||||||||||||||||||
Db    346 CCAAGGCCAACCGTGAGA 363


RESULT 12
US-09-960-352-12448
; Sequence 12448, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12448
;   LENGTH: 427
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 53-LIB34-010-Q1-E2-F10
US-09-960-352-12448

  Query Match             0.8%;  Score 18;  DB 10;  Length 427;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1160 CCAAGGCCAACCGTGAGA 1177
          ||||||||||||||||||
Db    335 CCAAGGCCAACCGTGAGA 352


RESULT 13
US-09-864-761-6022/c
; Sequence 6022, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6022
;   LENGTH: 466
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC004159.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
US-09-864-761-6022

  Query Match             0.8%;  Score 18;  DB 10;  Length 466;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    486 ACTGGCCAAAAAAGAAAA 503
          ||||||||||||||||||
Db    135 ACTGGCCAAAAAAGAAAA 118


RESULT 14
US-09-960-352-14538
; Sequence 14538, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
```

```
; APPLICANT:  Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE:  2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14538
;   LENGTH: 479
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 62-LIB3058-022-Q1-K1-H6
US-09-960-352-14538

  Query Match             0.8%;  Score 18;  DB 10;  Length 479;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1160 CCAAGGCCAACCGTGAGA 1177
          ||||||||||||||||||
Db    380 CCAAGGCCAACCGTGAGA 397


RESULT 15
US-09-918-995-1375
; Sequence 1375, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1375
;   LENGTH: 499
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(499)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-1375

  Query Match             0.8%;  Score 18;  DB 9;  Length 499;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    796 GACTATGCCATGACCATC 813
          ||||||||||||||||||
Db    327 GACTATGCCATGACCATC 344


Search completed: May 23, 2003, 09:58:00
Job time : 210 secs
```

D D

GenCore version 5.1.4_p5_4578


```
OM nucleic - nucleic search, using sw model

Run on:         May 23, 2003, 04:44:52 ; Search time 328 Seconds
                                          (without alignments)
                                          14686.050 Million cell updates/sec

Title:          US-10-008-355-1
Perfect score:  2139
Sequence:       1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:  OLIGO_NUC
                Gapop 60.0 , Gapext 60.0

Searched:       2185239 seqs, 1125999159 residues

Word size :      0

Total number of hits satisfying chosen parameters:      4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :       N_Geneseq_101002:*
                 1:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                 2:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                 3:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                 4:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                 5:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                 6:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                 7:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                 8:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                 9:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                 10:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                 11:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                 12:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                 13:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                 14:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                 15:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                 16:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                 17:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                 18:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                 19:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
                 20:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                 21:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                 22:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                 23:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                 24:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2139	100.0	2139	24	AAL43635	Porphyromonas ging
c	2	19	0.9	315	23	ABL27551	Drosophila melanog
	3	19	0.9	948	23	ABL03799	Drosophila melanog
	4	19	0.9	1431	24	ABQ70337	Listeria monocytog
	5	19	0.9	2252	23	ABL27548	Drosophila melanog
	6	19	0.9	2315	23	ABL27550	Drosophila melanog
	7	19	0.9	2324	23	ABL27552	Drosophila melanog
c	8	19	0.9	2916	23	ABL03838	Drosophila melanog
c	9	19	0.9	3006	23	ABL03798	Drosophila melanog
	10	19	0.9	4163	24	ABL65218	Lung cancer relate
c	11	19	0.9	4428	22	ABA09031	Human garp homolog
c	12	19	0.9	72928	20	AAZ18355	Human ASTHIJ 5' ge
c	13	19	0.9	72928	21	AAA80253	Human ASTH1J 5' ge
c	14	19	0.9	1830121	17	AAT42063	Haemophilus influe
	15	18	0.8	228	24	ABN69869	Streptococcus poly
c	16	18	0.8	252	24	ABK80475	Bacillus clausii g
	17	18	0.8	287	16	AAT23588	Human gene signatu
c	18	18	0.8	318	21	AAC28823	Human secreted pro
c	19	18	0.8	466	22	ABA58492	Human foetal liver
c	20	18	0.8	466	22	ABA27556	Probe #6022 for ge
c	21	18	0.8	466	22	AAK06605	Human brain expres
c	22	18	0.8	466	22	AAK32297	Human bone marrow
c	23	18	0.8	466	22	AAI16030	Probe #5963 for ge
c	24	18	0.8	466	22	AAI38149	Probe #6835 used t
c	25	18	0.8	466	24	ABS07075	Human genome-deriv
c	26	18	0.8	577	21	AAC54596	Arabidopsis thalia
c	27	18	0.8	579	21	AAC53594	Arabidopsis thalia
	28	18	0.8	650	24	ABQ59965	Human colon cancer
	29	18	0.8	756	22	AAH04348	Human cDNA clone (
	30	18	0.8	953	24	ABN74565	Bovine embryonic g
	31	18	0.8	1032	23	ABL06001	Drosophila melanog
	32	18	0.8	1251	18	AAV25022	H. pylori cell env
	33	18	0.8	1379	23	ABL07075	Drosophila melanog
c	34	18	0.8	1446	22	AAS15898	DNA encoding G-pro
	35	18	0.8	1490	21	AAC36486	Arabidopsis thalia
	36	18	0.8	1491	21	AAC49045	Arabidopsis thalia
	37	18	0.8	1908	18	AAV25132	H. pylori cell env
	38	18	0.8	1938	18	AAV24723	H. pylori ORF hplp
	39	18	0.8	2007	18	AAV25118	H. pylori cell env
	40	18	0.8	2070	21	AAZ88408	H. pylori bacteria
	41	18	0.8	2114	21	AAA87667	Human secreted pro
	42	18	0.8	2115	20	AAX75829	H. pylori outer me
	43	18	0.8	2127	21	AAZ88405	H. pylori bacteria
	44	18	0.8	2158	21	AAZ88409	H. pylori bacteria
	45	18	0.8	2226	21	AAZ88406	H. pylori bacteria

ALIGNMENTS

```
RESULT 1
AAL43635
ID   AAL43635 standard; DNA; 2139 BP.
XX
AC   AAL43635;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) coding sequence.
XX
KW   Dipeptidylpeptidase-7; DPP-7; gene; ds; enzyme; amidolytic cleavage;
KW   DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW   periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2139
FT                   /*tag=  a
FT                   /product= "Porphyromonas gingivalis DPP-7"
XX
PN   WO200238742-A2.
XX
PD   16-MAY-2002.
XX
PF   08-NOV-2001; 2001WO-US46782.
XX
PR   08-NOV-2000; 2000US-246827P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,  Potempa JS,  Banbula A,  Bugno M;
```

```
XX
DR   WPI; 2002-490075/52.
DR   P-PSDB; AAO15205.
XX
PT   Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT   the dipeptidylpeptidase for protecting an animal from periodontal
PT   disease caused by Porphyromonas gingivalis  -
XX
PS   Claim 11; Fig 4; 65pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC   enzyme has amidolytic activity for cleavage of a peptide bond between the
CC   second and third amino acids from the N-terminal end of a target peptide.
CC   The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC   substituent on the alpha-carbon atom of the second amino acid from the
CC   N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC   useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC   reducing the growth of a bacterium and protecting an animal from a
CC   periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC   or periodontitis). The present DNA sequence encodes the Porphyromonas
CC   gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the invention.
XX
SQ   Sequence 2139 BP; 544 A; 543 C; 565 G; 487 T; 0 other;

  Query Match             100.0%;  Score 2139;  DB 24;  Length 2139;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2139;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60

Qy         61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120

Qy        121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180

Qy        181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240

Qy        241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300

Qy        301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360

Qy        361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420

Qy        421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480

Qy        481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540

Qy        541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600

Qy        601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660

Qy        661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720

Qy        721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780

Qy        781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840

Qy        841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900

Qy        901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960

Qy        961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020

Qy       1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080

Qy       1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140

Qy       1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200

Qy       1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260

Qy       1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320

Qy       1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380

Qy       1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440

Qy       1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500

Qy       1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560

Qy       1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620

Qy       1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680

Qy       1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740

Qy       1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800

Qy       1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860

Qy       1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920

Qy       1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980

Qy       1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040

Qy       2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100

Qy       2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 2139
              |||||||||||||||||||||||||||||||||||||||
Db       2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 2139
```

```
RESULT 2
ABL27551/c
ID   ABL27551 standard; DNA; 315 BP.
XX
AC   ABL27551;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 34126.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 34126; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 315 BP; 68 A; 103 C; 83 G; 61 T; 0 other;

  Query Match             0.9%;  Score 19;  DB 23;  Length 315;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        262 CACCACTGCGGATACGGTG 280
              |||||||||||||||||||
Db        314 CACCACTGCGGATACGGTG 296


RESULT 3
ABL03799
ID   ABL03799 standard; cDNA; 948 BP.
XX
AC   ABL03799;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 5879.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59696.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 5879; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 948 BP; 377 A; 169 C; 195 G; 207 T; 0 other;

  Query Match             0.9%;  Score 19;  DB 23;  Length 948;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        485 AACTGGCCAAAAAAGAAAA 503
              |||||||||||||||||||
Db        721 AACTGGCCAAAAAAGAAAA 739
```

```
RESULT 4
ABQ70337
ID   ABQ70337 standard; DNA; 1431 BP.
XX
AC   ABQ70337;
XX
DT   29-AUG-2002  (first entry)
XX
DE   Listeria monocytogenes 4b contig DNA sequence #279.
XX
KW   Antibacterial; Listeria; food contamination; mutational analysis;
KW   infection; ds.
XX
OS   Listeria monocytogenes 4b.
XX
PN   WO200228891-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-FR03061.
XX
PR   04-OCT-2000; 2000FR-0012697.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Kunst F,  Glaser P;
XX
DR   WPI; 2002-332479/37.
XX
PT   New genomic sequences from Listeria species, useful for detection,
PT   treatment and prevention of infection, also related polypeptides,
PT   antibodies and modulators  -
XX
PS   Claim 14; SEQ ID 3150; 180pp; French.
XX
CC   The present invention relates to nucleic acid sequences
CC   (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC   and primers for identification and/or detection of Listeria (e.g. as
CC   contaminants in foods, or mutational analysis) and for analysis of
CC   gene expression. Proteins encoded by the nucleic acid sequences can be
CC   used to screen for compounds that modulate gene expression, replication
CC   and pathogenicity of Listeria (potential therapeutic agents), also for
CC   treating infections by Listeria, and are useful as immunogens in
CC   anti-Listeria vaccines.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1431 BP; 434 A; 278 C; 333 G; 386 T; 0 other;

  Query Match              0.9%;  Score 19;  DB 24;  Length 1431;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      490 GCCAAAAAAGAAAATGCAG 508
            |||||||||||||||||||
Db     1007 GCCAAAAAAGAAAATGCAG 1025


RESULT 5
ABL27548
ID   ABL27548 standard; DNA; 2252 BP.
XX
AC   ABL27548;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 34117.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 34117; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2252 BP; 652 A; 432 C; 535 G; 633 T; 0 other;

  Query Match              0.9%;  Score 19;  DB 23;  Length 2252;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      262 CACCACTGCGGATACGGTG 280
            |||||||||||||||||||
Db       14 CACCACTGCGGATACGGTG 32


RESULT 6
ABL27550
ID   ABL27550 standard; DNA; 2315 BP.
XX
AC   ABL27550;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 34123.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
```

```
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 34123; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2315 BP; 608 A; 443 C; 568 G; 696 T; 0 other;

  Query Match             0.9%;  Score 19;  DB 23;  Length 2315;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     262 CACCACTGCGGATACGGTG 280
           |||||||||||||||||||
Db    1002 CACCACTGCGGATACGGTG 1020


RESULT 7
ABL27552
ID   ABL27552 standard; DNA; 2324 BP.
XX
AC   ABL27552;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 34129.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 34129; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2324 BP; 672 A; 414 C; 466 G; 772 T; 0 other;

  Query Match             0.9%;  Score 19;  DB 23;  Length 2324;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     262 CACCACTGCGGATACGGTG 280
           |||||||||||||||||||
Db    2068 CACCACTGCGGATACGGTG 2086


RESULT 8
ABL03838/c
ID   ABL03838 standard; cDNA; 2916 BP.
XX
AC   ABL03838;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 5996.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59735.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 5996; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2916 BP; 739 A; 611 C; 476 G; 1090 T; 0 other;

  Query Match             0.9%;  Score 19;  DB 23;  Length 2916;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    485 AACTGGCCAAAAAAGAAAA 503
          |||||||||||||||||||
Db   2714 AACTGGCCAAAAAAGAAAA 2696


RESULT 9
ABL03798/c
ID   ABL03798 standard; cDNA; 3006 BP.
XX
AC   ABL03798;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 5876.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59695.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 5876; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3006 BP; 870 A; 568 C; 502 G; 1066 T; 0 other;

  Query Match             0.9%;  Score 19;  DB 23;  Length 3006;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    485 AACTGGCCAAAAAAGAAAA 503
          |||||||||||||||||||
Db   1228 AACTGGCCAAAAAAGAAAA 1210


RESULT 10
ABL65218
ID   ABL65218 standard; DNA; 4163 BP.
XX
AC   ABL65218;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Lung cancer related gene sequence SEQ ID NO:3555.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US10838.
XX
PR   05-JUN-2000; 2000US-209473P.
PR   05-JUN-2000; 2000US-209531P.
PR   18-SEP-2000; 2000US-233133P.
PR   18-SEP-2000; 2000US-233617P.
PR   20-SEP-2000; 2000US-234009P.
PR   20-SEP-2000; 2000US-234034P.
PR   20-SEP-2000; 2000US-234052P.
PR   22-SEP-2000; 2000US-234509P.
PR   22-SEP-2000; 2000US-234567P.
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set -
XX
PS   Claim 1; SEQ ID 3555; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
```

```
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 4163 BP; 836 A; 1303 C; 1112 G; 912 T; 0 other;

  Query Match             0.9%;  Score 19;  DB 24;  Length 4163;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        241 CAGGGCCTGATCTTTACCA 259
              |||||||||||||||||||
Db       3135 CAGGGCCTGATCTTTACCA 3153


RESULT 11
ABA09031/c
ID   ABA09031 standard; cDNA; 4428 BP.
XX
AC   ABA09031;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human garp homologue-encoding cDNA, SEQ ID NO:807.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157188-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03800.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457740/49.
DR   P-PSDB; ABB11787.
XX
PT   Human proteins and DNA encoding sequences useful for preventing,
PT   treating or ameliorating a medical condition in a mammalian subject
PT   e.g. arthritis and cancer -
XX
PS   Claim 1; Page 713-715; 1963pp; English.
XX
CC   Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine, cell proliferation or cell
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g., by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a cDNA encoding a
CC   novel human polypeptide of the invention.
XX
SQ   Sequence 4428 BP; 968 A; 1198 C; 1372 G; 890 T; 0 other;

  Query Match             0.9%;  Score 19;  DB 22;  Length 4428;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        241 CAGGGCCTGATCTTTACCA 259
              |||||||||||||||||||
Db       1033 CAGGGCCTGATCTTTACCA 1015


RESULT 12
AAZ18355/c
ID   AAZ18355 standard; DNA; 72928 BP.
XX
AC   AAZ18355;
XX
DT   19-OCT-1999  (first entry)
XX
DE   Human ASTHIJ 5' genomic region.
XX
KW   ASTH1; asthma; human; chromosome 11p; ASTH1I; ASTH1J; genetic locus;
KW   therapeutic; immunogen; ds.
XX
OS   Homo sapiens.
XX
PN   WO9937809-A1.
XX
PD   29-JUL-1999.
XX
PF   21-JAN-1998;   98WO-US01260.
XX
```

```
PR   21-JAN-1998;    98WO-US01260.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Brooks-Wilson AR,  Buckler A,  Cardon L,  Carey AH;
PI   Galvin M,  Miller A,  North M;
XX
DR   WPI; 1999-479058/40.
XX
PT   Mammalian asthma related genes, useful for diagnosis of a
PT   predisposition to development of asthma
XX
PS   Claim 15; Page 75-96; 195pp; English.
XX
CC   The invention identifies a genetic locus ASTH1, associated with asthma,
CC   mapped to human chromosome 11p. ASTH1I and ASTHIJ are genes present
CC   within the locus, located close to each other on human chromosome 11p,
CC   and have similar patterns of expression, and common sequence motifs. The
CC   ASTH1 genes and fragments, encoded protein, genomic regulatory regions
CC   and anti-ASTH1 antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for the modulation of gene
CC   activity in vivo for prophylactic and therapeutic purposes. The ASTH1
CC   protein is useful as an immunogen to raise specific antibodies, in drug
CC   screening for compositions that mimic or modulate ASTH1 activity or
CC   expression, including altered forms of ASTH1 protein, and as a
CC   therapeutic. The present sequence represents a human ASTH1 genomic region
CC   sequence.
XX
SQ   Sequence 72928 BP; 20042 A; 15100 C; 16036 G; 21750 T; 0 other;

  Query Match             0.9%;  Score 19;  DB 20;  Length 72928;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1625 CCATTGAGAAGGGCAAGCG 1643
          |||||||||||||||||||
Db  42822 CCATTGAGAAGGGCAAGCG 42804


RESULT 13
AAA80253/c
ID   AAA80253 standard; DNA; 72928 BP.
XX
AC   AAA80253;
XX
DT   22-NOV-2000  (first entry)
XX
DE   Human ASTH1J 5' genomic region.
XX
KW   ASTH1 locus; ASTH1J; human; chromosome 11p; asthma;
KW   bronchial hyperreactivity; ets family; transcription factor;
KW   splice variant; genetic predisposition; polymorphism; antibody;
KW   drug screening; prophylaxis; therapy; diagnosis; ds.
XX
OS   Homo sapiens.
XX
PN   US6087485-A.
XX
PD   11-JUL-2000.
XX
PF   21-JAN-1998;    98US-0009913.
XX
PR   21-JAN-1997;    97US-0035663.
PR   01-JUL-1997;    97US-0051432.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  North M,  Cardon L,  Buckler A;
PI.  Brooks-Wilson AR,  Carey AH;
XX
DR   WPI; 2000-505109/45.
XX
PT   New nucleic acids other than naturally occurring chromosomes encoding
PT   ASTH1 protein, for e.g. screening compositions that modulate expression
PT   or function of ASTH1 proteins or as diagnostics for genetic
PT   predisposition to asthma  -
XX
PS   Claim 7; Column 49-112; 131pp; English.
XX
CC   The invention relates to the ASTH1 locus on the short arm of human
CC   chromosome (11p). This locus comprises the ASTH1I and ASTH1J genes,
CC   which are associated with a genetic predisposition to asthma and
CC   bronchial hyperreactivity. The ASTH1I and ASTH1J genes are oriented in
CC   opposite directions with the ASTH1 locus, and have similar patterns of
CC   expression and common sequence motifs. They are both expressed in
CC   trachea, lung and several other tissues. ASTH1I and ASTH1J are novel
CC   members of the ets family of transcription factors, which have been
CC   implicated in the activation of a variety of genes including the TCRa
CC   gene and cytokine genes known to be important in the aetiology of asthma.
CC   Both ASTH1I and ASTH1J mRNAs are alternatively spliced. Alternative
CC   splicing of transcripts has no effect on the open reading frame of
CC   ASTH1J, as the exons involved are all 5' to the start codon in exon b. In
CC   contrast, alternative splicing of ASTH1I transcripts results in 3
CC   different ASTH1I isoforms. The invention also encompasses mouse asth1j
CC   protein. The ASTH1 nucleic acids are useful as diagnostics to identify a
CC   hereditary predisposition to asthma, as probes for identifying ASTH1
CC   related genes, for identifying expression of the gene in a biological
CC   specimen, and for generating genetically modified non-human animals or
CC   site specific gene modifications in cell lines. The encoded ASTH1
CC   proteins are useful as immunogens to raise specific antibodies; in drug
CC   screening for compositions that mimic or modulate activity or expression
CC   of ASTH1I and/or ASTH1J (including altered forms of these proteins); and
CC   as a therapeutic. The ASTH1 genes or fragments thereof, encoded proteins,
CC   ASTH1 genomic regulatory regions, and anti-ASTH1I and anti-ASTH1J
CC   antibodies are useful in the identification of individuals predisposed to
CC   development of asthma, and for modulation of gene activity in vivo for
CC   prophylactic and therapeutic purposes. The intact ASTH1I or ASTH1J
CC   proteins or active fragments thereof may be used to modulate or reduce
CC   bronchial hyperreactivity. The present sequence represents the
CC   non-transcribed region upstream of the human ASTH1J gene.
XX
SQ   Sequence 72928 BP; 20041 A; 15101 C; 16036 G; 21750 T; 0 other;

  Query Match             0.9%;  Score 19;  DB 21;  Length 72928;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1625 CCATTGAGAAGGGCAAGCG 1643
          |||||||||||||||||||
Db  42822 CCATTGAGAAGGGCAAGCG 42804


RESULT 14
AAT42063/c
ID   AAT42063 standard; DNA; 1830121 BP.
XX
AC   AAT42063;
XX
DT   14-SEP-1999  (first entry)
XX
DE   Haemophilus influenzae complete genome sequence.
XX
KW   Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW   expression modulating fragment; regulation; gene expression; vector;
KW   organism; open reading frame; ORF; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO9633276-A1.
XX
PD   24-OCT-1996.
XX
PF   22-APR-1996;    96WO-US05320.
XX
PR   07-JUN-1995;    95US-0487429.
PR   21-APR-1995;    95US-0426787.
```

```
PR   07-JUN-1995;   95US-0476102.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Adams MD,  Fleischmann RD,  Smith HO,  Venter JC,  White O;
XX
DR   WPI; 1996-485782/48.
XX
PT   Haemophilus influenzae Rd genome recorded on computer readable
PT   medium  - useful for identifying commercially important nucleic acid
PT   fragments by homology searching
XX
PS   Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC   This sequence represents the complete genome sequence of the bacterium
CC   Haemophilus influenzae strain Rd.  The invention relates to a computer
CC   readable medium (CRM) having recorded upon it the complete H.influenzae
CC   nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC   sequence at least 99% identical to (I).  By providing the full-length
CC   genomic sequence in a computer readable form, it is possible to identify
CC   commercially important nucleic acid fragments and expression modulating
CC   fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC   regulate the expression of a nucleic acid molecule.  Vectors and altered
CC   organisms comprising the predicted ORFS can be used to produce any of the
CC   polypeptide fragments of the H. influenzae Rd genome.
XX
SQ   Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

  Query Match             0.9%;  Score 19;  DB 17;  Length 1830121;
  Best Local Similarity  100.0%;  Pred. No. 29;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1138 GAAAAGGCTTATAAGGAAG 1156
           |||||||||||||||||||
Db 1428299 GAAAAGGCTTATAAGGAAG 1428281


RESULT 15
ABN69869
ID   ABN69869 standard; DNA; 228 BP.
XX
AC   ABN69869;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 7651.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus pyogenes.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB04789.
XX
PR   27-OCT-2000; 2000GB-0026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J,  Masignani V,  Margarit Ros YI,  Grandi G,  Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP29238.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein -
XX
PS   Claim 7; Page 3903; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence 228 BP; 85 A; 23 C; 46 G; 74 T; 0 other;

  Query Match             0.8%;  Score 18;  DB 24;  Length 228;
  Best Local Similarity  100.0%;  Pred. No. 80;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 ATGAAATTAAAAAGTATT 24
              ||||||||||||||||||
Db          4 ATGAAATTAAAAAGTATT 21


Search completed: May 23, 2003, 07:31:39
Job time : 2006 secs
```

```
                    GenCore version 5.1.4_p5_4578
              Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         May 23, 2003, 05:20:28 ; Search time 62 Seconds
                                         (without alignments)
                                         10580.348 Million cell updates/sec

Title:          US-10-008-355-1
Perfect score:  2139
Sequence:       1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:  OLIGO_NUC
                Gapop 60.0 , Gapext 60.0

Searched:       441362 seqs, 153338381 residues

Word size :     0

Total number of hits satisfying chosen parameters:      882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :       Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1500	70.1	1974	4	US-09-221-017B-726	Sequence 726, App
	2	92	4.3	561	4	US-09-221-017B-12	Sequence 12, Appl
c	3	19	0.9	72928	3	US-09-009-913-1	Sequence 1, Appli
	4	18	0.8	1317	4	US-09-221-017B-382	Sequence 382, App
	5	18	0.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	6	18	0.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	7	17	0.8	861	2	US-08-743-637B-166	Sequence 166, App
	8	17	0.8	861	3	US-08-526-840B-166	Sequence 166, App
c	9	17	0.8	1958	4	US-09-570-454-1	Sequence 1, Appli
c	10	17	0.8	2352	4	US-08-997-251-3	Sequence 3, Appli
c	11	17	0.8	2692	1	US-07-932-454A-2	Sequence 2, Appli
	12	17	0.8	2822	2	US-08-679-405-1	Sequence 1, Appli
	13	17	0.8	2822	2	US-08-842-799-1	Sequence 1, Appli
	14	17	0.8	2822	5	PCT-US96-11458-1	Sequence 1, Appli
	15	17	0.8	49272	1	US-08-614-770A-1	Sequence 1, Appli
c	16	16	0.7	50	2	US-08-832-468-5	Sequence 5, Appli
c	17	16	0.7	313	1	US-08-396-452-1	Sequence 1, Appli
c	18	16	0.7	313	4	US-09-169-119-1	Sequence 1, Appli
	19	16	0.7	360	4	US-09-060-756-211	Sequence 211, App
c	20	16	0.7	592	4	US-09-385-982-126	Sequence 126, App
c	21	16	0.7	615	4	US-08-998-416-1063	Sequence 1063, Ap
c	22	16	0.7	652	4	US-09-328-111-161	Sequence 161, App
c	23	16	0.7	705	4	US-08-998-416-1059	Sequence 1059, Ap
	24	16	0.7	735	4	US-09-535-008-32	Sequence 32, Appl
	25	16	0.7	1063	4	US-09-077-675A-1	Sequence 1, Appli
	26	16	0.7	1322	4	US-09-446-821A-2	Sequence 2, Appli
	27	16	0.7	1330	1	US-08-480-604A-22	Sequence 22, Appl
	28	16	0.7	1330	2	US-08-405-496A-22	Sequence 22, Appl
	29	16	0.7	1330	4	US-08-915-136-22	Sequence 22, Appl
	30	16	0.7	1347	2	US-08-959-749-1	Sequence 1, Appli
	31	16	0.7	1347	4	US-09-351-497-1	Sequence 1, Appli
	32	16	0.7	1359	4	US-09-134-001C-1336	Sequence 1336, Ap
	33	16	0.7	1402	1	US-08-480-604A-25	Sequence 25, Appl
	34	16	0.7	1402	2	US-08-405-496A-25	Sequence 25, Appl
	35	16	0.7	1402	4	US-08-915-136-25	Sequence 25, Appl
c	36	16	0.7	1491	4	US-08-740-223A-21	Sequence 21, Appl
c	37	16	0.7	1491	4	US-09-709-188-21	Sequence 21, Appl
c	38	16	0.7	1500	4	US-08-740-223A-23	Sequence 23, Appl
c	39	16	0.7	1500	4	US-09-709-188-23	Sequence 23, Appl
	40	16	0.7	1540	1	US-08-286-872-5	Sequence 5, Appli
c	41	16	0.7	1812	1	US-08-328-322-18	Sequence 18, Appl
c	42	16	0.7	1829	1	US-08-201-118-8	Sequence 8, Appli
c	43	16	0.7	1829	2	US-08-238-821B-8	Sequence 8, Appli
c	44	16	0.7	1829	5	PCT-US95-05744-8	Sequence 8, Appli
c	45	16	0.7	1964	1	US-08-328-322-9	Sequence 9, Appli

ALIGNMENTS

```
RESULT 1
US-09-221-017B-726
; Sequence 726, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  726:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1974 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
```

9/3/02

102e '98

1500

```
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:   1...1974
US-09-221-017B-726

  Query Match             70.1%;  Score 1500;  DB 4;  Length 1974;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1730;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    1;

Qy        409 GACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCT 468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCT 60

Qy        469 CAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATC 528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATC 120

Qy        529 GTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGAC 588
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGAC 180

Qy        589 GTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAAC 648
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAAC 240

Qy        649 TGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGAC 708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGAC 300

Qy        709 AACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCC 768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCC 360

Qy        769 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 420

Qy        829 ACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCT 888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCT 480

Qy        889 CGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAG 948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAG 540

Qy        949 GCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCG 1008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCG 600

Qy       1009 ATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAA 1068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAA 660

Qy       1069 AGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTG 1128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTG 720

Qy       1129 TCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTG 1188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTG 780

Qy       1189 AGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTG 1248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTG 840

Qy       1249 GCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGAC 1308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGAC 900

Qy       1309 TACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGT 1368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGT 960

Qy       1369 ATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC 1428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC 1020

Qy       1429 GACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAG 1488
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAG 1080

Qy       1489 TTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGAT 1548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGAT 1140

Qy       1549 CCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCG 1608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCG 1200

Qy       1609 ATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAG 1668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAG 1260

Qy       1669 ATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGC 1728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGC 1320

Qy       1729 TCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAG 1788
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAG 1380

Qy       1789 GGCGTATTGGAGAAGCAGGAT--CCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCC 1846
              ||||||||||||||||||||   |  ||||||||||||||||||||||||||||||||||
Db       1381 GGCGTATTGGAGAAGCAGGATTCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCC 1440

Qy       1847 TCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCG 1906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCG 1500

Qy       1907 CTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGA 1966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGA 1560

Qy       1967 ACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCG 2026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCG 1620

Qy       2027 AGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGA 2086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGA 1680

Qy       2087 TTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 2139
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 1733


RESULT 2
US-09-221-017B-12
; Sequence 12, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...561
US-09-221-017B-12

  Query Match             4.3%;  Score 92;  DB 4;  Length 561;
  Best Local Similarity   99.3%;  Pred. No. 1e-38;
  Matches  142;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     366 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTAAGGG 425

Qy      61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     426 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 485

Qy     121 CGAATGCGTGAGCTCGGCTTTAC 143
             |||||||||||||||||||||||
Db     486 CGAATGCGTGAGCTCGGCTTTAC 508


RESULT 3
US-09-009-913-1/c
; Sequence 1, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;    APPLICANT:  AxyS Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  Asthma Related Genes
;    NUMBER OF SEQUENCES:  339
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/009,913
;      FILING DATE:  21-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  SEQ-4P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3231
;      TELEFAX:  650-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  72928 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-09-009-913-1

  Query Match             0.9%;  Score 19;  DB 3;  Length 72928;
  Best Local Similarity   100.0%;  Pred. No. 4.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1625 CCATTGAGAAGGGCAAGCG 1643
           |||||||||||||||||||
Db   42822 CCATTGAGAAGGGCAAGCG 42804


RESULT 4
US-09-221-017B-382
; Sequence 382, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  382:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1317 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:   1...1317
US-09-221-017B-382

  Query Match             0.8%;  Score 18;  DB 4;  Length 1317;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     636 CGATACGGACAACTGGAT 653
           ||||||||||||||||||
Db     549 CGATACGGACAACTGGAT 566


RESULT 5
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             0.8%;  Score 18;  DB 4;  Length 4403765;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1325 ACCGTAAGGTGCTGCCCG 1342
           ||||||||||||||||||
Db 735985 ACCGTAAGGTGCTGCCCG 736002


RESULT 6
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             0.8%;  Score 18;  DB 4;  Length 4411529;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1325 ACCGTAAGGTGCTGCCCG 1342
           ||||||||||||||||||
Db 734553 ACCGTAAGGTGCTGCCCG 734570


RESULT 7
US-08-743-637B-166
; Sequence 166, Application US/08743637B
; Patent No. 5994066
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  PICARD, Francois J.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIES-SPECIFIC AND UNIVERSAL DNA
;    TITLE OF INVENTION:  PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
;    TITLE OF INVENTION:  IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
;    TITLE OF INVENTION:  ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
;    NUMBER OF SEQUENCES:  273
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 EAST WISCONSIN AVENUE
;      CITY:  MILWAUKEE
;      STATE:  WISCONSIN
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/743,637B
;      FILING DATE:  04-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/526,840
;      FILING DATE:  11-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414)277-5591
;  INFORMATION FOR SEQ ID NO:  166:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  861 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-743-637B-166

  Query Match              0.8%;  Score 17;  DB 2;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       37 GCCCTGCTGTTGGGTGC 53
            |||||||||||||||||
Db      466 GCCCTGCTGTTGGGTGC 482

RESULT 8
US-08-526-840B-166
; Sequence 166, Application US/08526840B
; Patent No. 6001564
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIFIC AND UNIVERSAL PROBES AND
;    TITLE OF INVENTION:  AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
;    TITLE OF INVENTION:  COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
;    TITLE OF INVENTION:  FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
;    NUMBER OF SEQUENCES:  177
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/526,840B
;      FILING DATE:  11-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/304,732
;      FILING DATE:  12-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414) 277-5591
;  INFORMATION FOR SEQ ID NO:  166:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  861 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-526-840B-166

  Query Match              0.8%;  Score 17;  DB 3;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       37 GCCCTGCTGTTGGGTGC 53
            |||||||||||||||||
Db      466 GCCCTGCTGTTGGGTGC 482

RESULT 9
US-09-570-454-1/c
; Sequence 1, Application US/09570454
; Patent No. 6399743
; GENERAL INFORMATION:
;  APPLICANT:  Department of Veterans Affairs
;  TITLE OF INVENTION:  Isolation and charaterization of epidermal growth
;  TITLE OF INVENTION:  factor releted protein
;  FILE REFERENCE:  107999.00106
;  CURRENT APPLICATION NUMBER: US/09/570,454
;  CURRENT FILING DATE:  2000-05-12
;  PRIOR APPLICATION NUMBER:  60/134,200
;  PRIOR FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS:  8
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1958
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-570-454-1

  Query Match              0.8%;  Score 17;  DB 4;  Length 1958;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1613 CCAATGCCTATGCCATT 1629
            |||||||||||||||||
Db     1255 CCAATGCCTATGCCATT 1239

RESULT 10
US-08-997-251-3/c
; Sequence 3, Application US/08997251
; Patent No. 6271440
;  GENERAL INFORMATION:
;    APPLICANT:  GUBLER, FRANZ J.
;    APPLICANT:  JACOBSEN, JOHN V.
;    TITLE OF INVENTION:  PLANT REGULATORY PROTEINS III
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Boulevard
;      CITY:  Boulder
;      STATE:  CO
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/997,251
;      FILING DATE:  23-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO AU96/00383
;      FILING DATE:  21-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PN6470/95
;      FILING DATE:  09-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PN3779/95
;      FILING DATE:  23-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Winner, Ellen P.
;      REGISTRATION NUMBER:  28,547
;      REFERENCE/DOCKET NUMBER:  110-97
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  3:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2352 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Oryza sativa
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  396..2054
US-08-997-251-3

  Query Match              0.8%;  Score 17;  DB 4;  Length 2352;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      601 TTTGCTCCTCCCAGCTC 617
            |||||||||||||||||
Db      309 TTTGCTCCTCCCAGCTC 293


RESULT 11
US-07-932-454A-2/c
; Sequence 2, Application US/07932454A
; Patent No. 5262318
;  GENERAL INFORMATION:
;    APPLICANT:  GUTHRIE, ELLEN P.
;    TITLE OF INVENTION:  ISOLATED DNA ENCODING THE SphI
;    TITLE OF INVENTION:  RESTRICTION ENDONUCLEASE AND RELATED METHODS FOR PRODUCING
;    TITLE OF INVENTION:  THE SAME
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 WATER STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/932,454A
;      FILING DATE:  19920820
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WILLIAMS, GREGORY D.
;      REGISTRATION NUMBER:  30901
;      REFERENCE/DOCKET NUMBER:  42078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 523-3400
;      TELEFAX:  (617) 5523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2692 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  703..1653
;      OTHER INFORMATION:  /note= "METHYLASE GENE STARTS AT
;      OTHER INFORMATION:  POSITION 703/ENDS AT 1653. 'RESTRICTION
;      OTHER INFORMATION:  ENDONUCLEASE STARTS AT POSITION 1703/ENDS AT 2410"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1703..2410
US-07-932-454A-2

  Query Match              0.8%;  Score 17;  DB 1;  Length 2692;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1394 TATTCAAGAATGTAATC 1410
            |||||||||||||||||
Db     1809 TATTCAAGAATGTAATC 1793


RESULT 12
US-08-679-405-1
; Sequence 1, Application US/08679405
; Patent No. 5866393
;  GENERAL INFORMATION:
;    APPLICANT:  Fuglsang, Claus
;    APPLICANT:  Halkier, Torben
;    APPLICANT:  Oxenboll, Karen M.
;    APPLICANT:  Berka, Randy M.
;    APPLICANT:  Cherry, Joel
;    TITLE OF INVENTION:  Haloperoxidases from Curvularia
;    TITLE OF INVENTION:  Verruculosa and Nucleic Acids Encoding Same
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5866393o No. 5866393disk of No. 5866393th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/679,405
;      FILING DATE:  July 9, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/001,194
;      FILING DATE:  July 14, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/603,534
;      FILING DATE:  February 21, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4441.210.US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 867-0123
;      TELEFAX:  (212) 878-9655
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  477..2276
US-08-679-405-1

  Query Match              0.8%;  Score 17;  DB 2;  Length 2822;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 GCAAATGAAATTAAAAA 19
            |||||||||||||||||
Db     2629 GCAAATGAAATTAAAAA 2645
```

```
RESULT 13
US-08-842-799-1
; Sequence 1, Application US/08842799
; Patent No. 5965418
;  GENERAL INFORMATION:
;    APPLICANT:  Fuglsang, Claus
;    APPLICANT:  Halkier, Torben
;    APPLICANT:  Oxenboll, Karen M.
;    APPLICANT:  Berka, Randy M.
;    APPLICANT:  Cherry, Joel
;    TITLE OF INVENTION:  Haloperoxidases from Curvularia
;    TITLE OF INVENTION:  Verruculosa and Nucleic Acids Encoding Same
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5965418o No. 5965418disk of No. 5965418th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/842,799
;      FILING DATE:  16-APR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/679,405
;      FILING DATE:  July 9, 1996
;      APPLICATION NUMBER:  60/001,194
;      FILING DATE:  July 14, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/603,534
;      FILING DATE:  February 21, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4441.210.US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 867-0123
;      TELEFAX:  (212) 878-9655
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  477..2276
US-08-842-799-1

  Query Match              0.8%;  Score 17;  DB 2;  Length 2822;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 GCAAATGAAATTAAAAA 19
            |||||||||||||||||
Db     2629 GCAAATGAAATTAAAAA 2645


RESULT 14
PCT-US96-11458-1
; Sequence 1, Application PC/TUS9611458
;  GENERAL INFORMATION:
;    APPLICANT:
;    APPLICANT:
;    TITLE OF INVENTION:  Haloperoxidases from Curvularia
;    TITLE OF INVENTION:  Verruculosa and Nucleic Acids Encoding Same
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Novo Nordisk of North America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/11458
;      FILING DATE:  9-JUL-1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/001,194
;      FILING DATE:  14-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/603,534
;      FILING DATE:  21-FEB-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4441.204-WO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 867-0123
;      TELEFAX:  (212) 878-9655
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2822 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  477..2276
PCT-US96-11458-1

  Query Match              0.8%;  Score 17;  DB 5;  Length 2822;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 GCAAATGAAATTAAAAA 19
            |||||||||||||||||
Db     2629 GCAAATGAAATTAAAAA 2645


RESULT 15
US-08-614-770A-1
; Sequence 1, Application US/08614770A
; Patent No. 5773267
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAM R. JACOBS AND GRAHAM F. HATFULL
;    TITLE OF INVENTION:  D29 SHUTTLE PHASMIDS AND USES THEREOF
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  AMSTER, ROTHSTEIN & EBENSTEIN
;      STREET:  90 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH 1.44 Mb STORAGE
;      MEDIUM TYPE:  DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/614,770A
;      FILING DATE:  MARCH 7, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ELIZABETH A. BOGOSIAN
;      REGISTRATION NUMBER:  39,911
;      REFERENCE/DOCKET NUMBER:  96700/402
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49272
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:
;      DESCRIPTION:  OLIGONUCLEOTIDE
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MYCOBACTERIOPHAGE
;      INDIVIDUAL ISOLATE:  D29
US-08-614-770A-1

  Query Match             0.8%;  Score 17;  DB 1;  Length 49272;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    836 GCTACCTCACTTCTTGG 852
          |||||||||||||||||
Db  46900 GCTACCTCACTTCTTGG 46916

Search completed: May 23, 2003, 10:27:15
Job time : 6285 secs
```

DD

```
                    GenCore version 5.1.4_p5_4578
             Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          May 23, 2003, 05:02:37 ; Search time 2048 Seconds
                                         (without alignments)
                                         16915.110 Million cell updates/sec

Title:           US-10-008-355-1
Perfect score:   2139
Sequence:        1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:   OLIGO_NUC
                 Gapop 60.0 , Gapext 60.0

Searched:        16154066 seqs, 8097743376 residues

Word size :       0

Total number of hits satisfying chosen parameters:      32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        EST:*
                 1:  em_estba:*
                 2:  em_esthum:*
                 3:  em_estin:*
                 4:  em_estmu:*
                 5:  em_estov:*
                 6:  em_estpl:*
                 7:  em_estro:*
                 8:  em_htc:*
                 9:  gb_est1:*
                 10:  gb_est2:*
                 11:  gb_htc:*
                 12:  gb_est3:*
                 13:  gb_est4:*
                 14:  gb_est5:*
                 15:  em_estfun:*
                 16:  em_estom:*
                 17:  gb_gss:*
                 18:  em_gss_hum:*
                 19:  em_gss_inv:*
                 20:  em_gss_pln:*
                 21:  em_gss_vrt:*
                 22:  em_gss_fun:*
                 23:  em_gss_mam:*
                 24:  em_gss_mus:*
                 25:  em_gss_other:*
                 26:  em_gss_pro:*
                 27:  em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21	1.0	444	17	AQ820052	AQ820052 HS_5270_A
	2	21	1.0	591	9	AI928924	AI928924 au58a07.y
	3	21	1.0	800	17	AG084100	AG084100 Pan trogl
c	4	21	1.0	1061	13	BM553311	BM553311 AGENCOURT
	5	20	0.9	223	10	AW200173	AW200173 da12e08.y
	6	20	0.9	257	10	AW881153	AW881153 QV0-OT003
c	7	20	0.9	260	12	BG158262	BG158262 EM1_9_F06
c	8	20	0.9	410	10	AW199497	AW199497 da12e08.x
	9	20	0.9	419	12	BE809917	BE809917 217225 MA
	10	20	0.9	463	14	BQ412487	BQ412487 GA__Ed005
c	11	20	0.9	472	14	BQ403470	BQ403470 GA__Ed005
	12	20	0.9	473	12	BF654963	BF654963 279329 MA
c	13	20	0.9	509	9	AI718392	AI718392 as59g06.x
	14	20	0.9	519	12	BE809904	BE809904 217039 MA
	15	20	0.9	523	13	BI847597	BI847597 469747 MA
c	16	20	0.9	531	13	BJ080132	BJ080132 BJ080132
	17	20	0.9	547	10	AW653212	AW653212 101812 MA
c	18	20	0.9	568	17	CNS07HUQ	AL611540 Anopheles
c	19	20	0.9	570	9	AA582263	AA582263 nn52h05.s
c	20	20	0.9	668	13	BJ043920	BJ043920 BJ043920
	21	20	0.9	787	13	BM029390	BM029390 IpSkn0090
	22	20	0.9	988	13	BM321169	BM321169 rockefell
c	23	19	0.9	313	14	D68316	D68316 CELK129H2F
c	24	19	0.9	334	10	AW597013	AW597013 sj66c04.y
	25	19	0.9	352	17	AQ057720	AQ057720 CIT-HSP-2
	26	19	0.9	358	14	T68361	T68361 yc41c12.r1
c	27	19	0.9	360	14	D68463	D68463 CELK131F7F
c	28	19	0.9	360	14	D69805	D69805 CELK092DXF
	29	19	0.9	368	14	BQ301839	BQ301839 RC1-BT025
	30	19	0.9	368	14	BQ301843	BQ301843 RC1-BT025
c	31	19	0.9	369	17	BH807962	BH807962 1008074D0
	32	19	0.9	376	10	BB809399	BB809399 BB809399
	33	19	0.9	402	9	AA062801	AA062801 zm04h07.r
c	34	19	0.9	414	17	AZ841761	AZ841761 2M0139M22
	35	19	0.9	418	14	R50228	R50228 yj61g03.r1
	36	19	0.9	448	14	H73815	H73815 ys11a08.r1
	37	19	0.9	462	14	T71455	T71455 yc61d05.r1
c	38	19	0.9	498	13	BJ107698	BJ107698 BJ107698
c	39	19	0.9	510	17	BH412712	BH412712 1007028G0
	40	19	0.9	522	17	AQ791029	AQ791029 HS_2239_B
	41	19	0.9	523	17	AQ844615	AQ844615 an42g06 J
c	42	19	0.9	524	17	BH801026	BH801026 1008108H1
	43	19	0.9	565	9	AI606604	AI606604 mz99h03.y
c	44	19	0.9	587	13	BM504607	BM504607 ig88c06.x
c	45	19	0.9	599	9	AU205200	AU205200 AU205200

ALIGNMENTS

```
RESULT 1
AQ820052
LOCUS       AQ820052                 444 bp    DNA     linear   GSS 26-AUG-1999
DEFINITION  HS_5270_A1_H10_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=846 Col=19 Row=O, DNA sequence.
ACCESSION   AQ820052
VERSION     AQ820052.1  GI:5782445
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 444)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
```

20 '99

```
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 846  row: O  column: 19
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 444.
FEATURES             Location/Qualifiers
     source          1. .444
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=846 Col=19 Row=O"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      101 a    117 c    113 g    111 t      2 others
ORIGIN

  Query Match             1.0%;  Score 21;  DB 17;  Length 444;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     338 GCACGATGGGTGAGGAGCTTC 358
           |||||||||||||||||||||
Db     230 GCACGATGGGTGAGGAGCTTC 250


RESULT 2
AI928924
LOCUS       AI928924                 591 bp    mRNA    linear   EST 23-AUG-1999
DEFINITION  au58a07.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone
            IMAGE:2518932 5' similar to gb:X63432_cds1 ACTIN, CYTOPLASMIC 1
            (HUMAN);, mRNA sequence.
ACCESSION   AI928924
VERSION     AI928924.1  GI:5664888
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 329.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2518932"
                     /clone_lib="Schneider fetal brain 00004"
                     /sex="male"
                     /tissue_type="frontal lobe"
                     /dev_stage="5 months post-conception"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescript SK (Stratagene);
                     Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
                     prepared from human fetal brain tissue.  5' and 3'
                     adaptors were used in cloning as follows: 5' adaptor
                     sequence:
                     5'-GAGAGAGAGAAGAGCTCAAGGATCCTTAATTAAATTAATCCCCCCCCCCCC-3'
                     and 3' adaptor sequence:
                     5'-GAGAGAGAGACTCGAGTTTTTTTTTTTTTTTTTT-3'.  The library was
                     size-selected for >0.5 kb inserts and has an average
                     insert size estimated at 1.2 kb.  This library was
                     constructed using the CAP-trapper method for full-length
                     enrichment and has not undergone amplification.   Library
                     was constructed by Dr. Claudio Schneider (LNCIB-Area
                     Science Park,  Trieste, Italy). "
BASE COUNT      115 a    207 c    164 g    104 t      1 others
ORIGIN

  Query Match             1.0%;  Score 21;  DB 9;  Length 591;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps  . 0;

Qy    1160 CCAAGGCCAACCGTGAGATGA 1180
           |||||||||||||||||||||
Db     409 CCAAGGCCAACCGTGAGATGA 429


RESULT 3
AG084100
LOCUS       AG084100                 800 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-081L22.R, genomic survey sequence.
ACCESSION   AG084100
VERSION     AG084100.1  GI:16635902
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-081L22.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 800)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector     : pKS145
                 R.Site 1   : SacI
                 R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-081L22.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      173 a    214 c    148 g    254 t     11 others
ORIGIN

  Query Match             1.0%;  Score 21;  DB 17;  Length 800;
  Best Local Similarity   100.0%;  Pred. No. 22;
```

```
  Matches    21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      553 GAATACTTCCTCATCGTCTAC 573
            |||||||||||||||||||||
Db      575 GAATACTTCCTCATCGTCTAC 595


RESULT 4
BM553311/c
LOCUS       BM553311               1061 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  AGENCOURT_6542482 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742946
            5', mRNA sequence.
ACCESSION   BM553311
VERSION     BM553311.1  GI:18791932
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1061)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12761  row: o  column: 11
            High quality sequence stop: 682.
FEATURES             Location/Qualifiers
     source          1..1061
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5742946"
                     /clone_lib="NIH_MGC_119"
                     /tissue_type="medulla"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source normal medulla from
                     anonymous male age 27.  Library  is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.3 kb, insert size range
                     0.9-3 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 013. Note:
                     this is a NIH_MGC Library."
BASE COUNT      254 a    310 c    287 g    209 t      1 others
ORIGIN

  Query Match             1.0%;  Score 21;  DB 13;  Length 1061;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       13 TTAAAAAGTATTCTTCTCGGA 33
            |||||||||||||||||||||
Db      897 TTAAAAAGTATTCTTCTCGGA 877


RESULT 5
AW200173
LOCUS       AW200173                223 bp    mRNA    linear   EST 27-JUN-2001
DEFINITION  da12e08.y1 Xenopus laevis oocyte Xenopus laevis cDNA clone
            XENOPUS_SOURCE_ID:xlnoc003o15 5' similar to TR:O35934 O35934 CLN3.
            ;, mRNA sequence.
ACCESSION   AW200173
VERSION     AW200173.1  GI:6480742
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 223)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: da12e08.x1
            Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Bruce Blumberg
             Library normalized by Jihwan Song
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: Xenopus clone distribution information for
            this library can be found through Research Genetics, visit their
            web page at: http://www.resgen.com/
            Seq primer: -40RP from Gibco
            High quality sequence stop: 125.
FEATURES             Location/Qualifiers
     source          1..223
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XENOPUS_SOURCE_ID:xlnoc003o15"
                     /clone_lib="Xenopus laevis oocyte"
                     /tissue_type="oocyte (stages 5 and 6)"
                     /lab_host="Top-10 F'"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from 2ug of poly A+ RNA.
                     EcoRI-XhoI cut cDNA was then ligated into UniZap-XR
                     (Stratagene) with EcoRI at the 5' end and XhoI at the 3'
                     end. SS-library phagemids were prepared by mass excision
                     from the original library and normalized by hybridization
                     to biotinylated driver (prepared from the same library by
                     PCR) to Cot-omega of 11. After removal of hybrids and
                     excess driver by streptavidin sepharose chromatography,
                     the ss-phagemids were made double stranded and
                     electroporated into Top-10 F'. Original library
                     contruction by Bruce Blumberg (Blumberg et al., 1991
                     Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,
                     2923-2935). Normalized by Jihwan Song (Song, Cho and
                     Blumberg, unpublished). Note: This is a Xenopus Gene
                     Collection (XGC) library."
BASE COUNT       39 a     60 c     56 g     68 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 10;  Length 223;
  Best Local Similarity   100.0%;  Pred. No. 59;
  Matches    20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      102 CAATCAGGAGAATCTGGATC 121
            ||||||||||||||||||||
Db      159 CAATCAGGAGAATCTGGATC 178


RESULT 6
AW881153
LOCUS       AW881153                257 bp    mRNA    linear   EST 23-MAY-2000
DEFINITION  QV0-OT0033-170400-197-c10 OT0033 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW881153
VERSION     AW881153.1  GI:8043163
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 257)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV0-OT0033-170
            400-197-c10&t3=2000-04-17&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 257.
FEATURES             Location/Qualifiers
     source          1. .257
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="OT0033"
                     /dev_stage="Adult"
                     /note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       86 a     30 c     57 g     84 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 10;  Length 257;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     956 GTATCAAATATGCCAGCAAG 975
           ||||||||||||||||||||
Db      83 GTATCAAATATGCCAGCAAG 102


RESULT 7
BG158262/c
LOCUS       BG158262                 260 bp    mRNA    linear   EST 08-MAR-2001
DEFINITION  EM1_9_F06.b1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG158262
VERSION     BG158262.1  GI:12691912
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 260)
  AUTHORS   Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
  TITLE     An EST database from Sorghum: developing embryos
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence start: 5
            High quality sequence stop: 116
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .260
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Embryo 1 (EM1)"
                     /note="Organ: Embryos germinated for 24 hr; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       71 a     69 c    110 g     10 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 12;  Length 260;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1020 CCGCGGTCTCGCTCGTCTTG 1039
           ||||||||||||||||||||
Db     160 CCGCGGTCTCGCTCGTCTTG 141


RESULT 8
AW199497/c
LOCUS       AW199497                 410 bp    mRNA    linear   EST 27-JUN-2001
DEFINITION  da12e08.x1 Xenopus laevis oocyte Xenopus laevis cDNA clone
            XENOPUS_SOURCE_ID:xlnoc003o15 3' similar to SW:CLN3_CANFA Q29611
            CLN3 PROTEIN. ;, mRNA sequence.
ACCESSION   AW199497
VERSION     AW199497.1  GI:6480066
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Bruce Blumberg
             Library normalized by Jihwan Song
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: Xenopus clone distribution information for
            this library can be found through Research Genetics, visit their
            web page at: http://www.resgen.com/
            Seq primer: -40UP from Gibco
            High quality sequence stop: 286.
FEATURES             Location/Qualifiers
     source          1. .410
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XENOPUS_SOURCE_ID:xlnoc003o15"
                     /clone_lib="Xenopus laevis oocyte"
```

```
                    /tissue_type="oocyte (stages 5 and 6)"
                    /lab_host="Top-10 F'"
                    /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                    XhoI; cDNA was prepared from 2ug of poly A+ RNA.
                    EcoRI-XhoI cut cDNA was then ligated into UniZap-XR
                    (Stratagene) with EcoRI at the 5' end and XhoI at the 3'
                    end. SS-library phagemids were prepared by mass excision
                    from the original library and normalized by hybridization
                    to biotinylated driver (prepared from the same library by
                    PCR) to Cot-omega of 11. After removal of hybrids and
                    excess driver by streptavidin sepharose chromatography,
                    the ss-phagemids were made double stranded and
                    electroporated into Top-10 F'. Original library
                    contruction by Bruce Blumberg (Blumberg et al., 1991
                    Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,
                    2923-2935). Normalized by Jihwan Song (Song, Cho and
                    Blumberg, unpublished). Note: This is a Xenopus Gene-
                    Collection (XGC) library."
BASE COUNT      121 a    101 c    112 g     76 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 10;  Length 410;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        102 CAATCAGGAGAATCTGGATC 121
              ||||||||||||||||||||
Db        373 CAATCAGGAGAATCTGGATC 354


RESULT 9
BE809917
LOCUS       BE809917                 419 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  217225 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE809917
VERSION     BE809917.1  GI:10241029
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 75  row: G  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..419
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT       76 a    147 c    118 g     78 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 12;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 67;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         24 TCTTCTCGGAGCAGCCCTGC 43
              ||||||||||||||||||||
Db         95 TCTTCTCGGAGCAGCCCTGC 114


RESULT 10
BQ412487
LOCUS       BQ412487                 463 bp    mRNA    linear   EST 22-MAY-2002
DEFINITION  GA__Ed0058F03r Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ed0058F03r, mRNA sequence.
ACCESSION   BQ412487
VERSION     BQ412487.1  GI:21100174
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total High Quality bases = 447
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence start: 2
            High quality sequence stop: 464.
FEATURES             Location/Qualifiers
     source          1..463
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ed0058F03r"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      114 a    144 c     65 g    140 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 14;  Length 463;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1489 TTCCATGCCATGCTCAAGTC 1508
              ||||||||||||||||||||
Db         87 TTCCATGCCATGCTCAAGTC 106


RESULT 11
BQ403470/c
LOCUS       BQ403470                 472 bp    mRNA    linear   EST 22-MAY-2002
```

```
DEFINITION  GA__Ed0058F03f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ed0058F03f, mRNA sequence.
ACCESSION   BQ403470
VERSION     BQ403470.1  GI:21091157
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 472)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total High Quality bases = 216
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence start: 3
            High quality sequence stop: 446.
FEATURES             Location/Qualifiers
     source          1. .472
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ed0058F03f"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      137 a     80 c    131 g    124 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 14;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1489 TTCCATGCCATGCTCAAGTC 1508
           ||||||||||||||||||||
Db     378 TTCCATGCCATGCTCAAGTC 359


RESULT 12
BF654963
LOCUS       BF654963                 473 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  279329 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF654963
VERSION     BF654963.1  GI:11920095
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 473)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 74  row: G  column: 8
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .473
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       82 a    175 c    125 g     91 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 12;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      24 TCTTCTCGGAGCAGCCCTGC 43
           ||||||||||||||||||||
Db     383 TCTTCTCGGAGCAGCCCTGC 402


RESULT 13
AI718392/c
LOCUS       AI718392                 509 bp    mRNA    linear   EST 10-JUN-1999
DEFINITION  as59g06.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone
            IMAGE:2333050 3' similar to SW:Y281_HUMAN Q92556 HYPOTHETICAL
            PROTEIN KIAA0281 ;, mRNA sequence.
ACCESSION   AI718392
VERSION     AI718392.1  GI:5035648
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40UP from Gibco
            High quality sequence stop: 220.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2333050"
```

```
                /clone_lib="Barstead colon HPLRB7"
                /sex="male"
                /dev_stage="adult, age 25"
                /lab_host="DH10B (phage resistant)"
                /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
                strand cDNA was primed with a Not I - oligo(dT) primer [5'
                TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                3']; double-stranded cDNA was ligated to Eco RI adaptors
                [5' AATTCACTAGTAAT 3' and 5' ATTACTAGTG 3'], digested
                with Not I and cloned into the Not I and Eco RI sites of
                the modified pT7T3 vector. Library constructed by Bob
                Barstead."
BASE COUNT      114 a    123 c    140 g    132 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 9;  Length 509;
  Best Local Similarity   100.0%;  Pred. No. 69;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   2089 GACAAATGGGGTCAGTGCCC 2108
          ||||||||||||||||||||
Db    369 GACAAATGGGGTCAGTGCCC 350


RESULT 14
BE809904
LOCUS       BE809904                 519 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  217039 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE809904
VERSION     BE809904.1  GI:10241016
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 75  row: D  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..519
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT       85 a    181 c    146 g    107 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 12;  Length 519;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     24 TCTTCTCGGAGCAGCCCTGC 43
          ||||||||||||||||||||
Db     95 TCTTCTCGGAGCAGCCCTGC 114


RESULT 15
BI847597
LOCUS       BI847597                 523 bp    mRNA    linear   EST 04-OCT-2001
DEFINITION  469747 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BI847597
VERSION     BI847597.1  GI:15960116
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 94  row: D  column: 18
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..523
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT       86 a    180 c    150 g    107 t
ORIGIN

  Query Match             0.9%;  Score 20;  DB 13;  Length 523;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     24 TCTTCTCGGAGCAGCCCTGC 43
          ||||||||||||||||||||
Db     96 TCTTCTCGGAGCAGCCCTGC 115


Search completed: May 23, 2003, 08:42:26
Job time : 2065 secs
```

D P

```
                GenCore version 5.1.4_p5_4578
           Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:         May 16, 2003, 13:03:17 ; Search time 30 Seconds
                                          (without alignments)
                                          4195.709 Million cell updates/sec

Title:          US-10-008-355-1
Perfect score:  712
Sequence:       1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:  OLIGO
                Xgapop 60.0 , Xgapext 60.0
                Ygapop 60.0 , Ygapext 60.0
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       262574 seqs, 29422922 residues

Word size:      1

Total number of hits satisfying chosen parameters:      438908

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110427_9763/app_query.fasta_1.2311
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=olin2p.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_33_@runat_16052003_110427_9763 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :      Issued_Patents_AA:*
                1:  /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
                2:  /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
                3:  /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
                4:  /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
                5:  /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                6:  /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	10	1.4	213	3	US-08-523-373-22	Sequence 22, Appl
2	10	1.4	214	3	US-08-523-373-23	Sequence 23, Appl
3	10	1.4	215	3	US-08-523-373-24	Sequence 24, Appl
4	10	1.4	344	1	US-08-657-192-3	Sequence 3, Appli
5	10	1.4	344	3	US-08-523-373-5	Sequence 5, Appli
6	10	1.4	392	3	US-08-523-373-6	Sequence 6, Appli
7	10	1.4	532	1	US-08-657-192-9	Sequence 9, Appli
8	10	1.4	532	3	US-08-523-373-7	Sequence 7, Appli
9	10	1.4	537	1	US-08-657-192-15	Sequence 15, Appl
10	9	1.3	15	2	US-08-332-562A-67	Sequence 67, Appl
11	9	1.3	311	4	US-09-134-001C-3729	Sequence 3729, Ap
12	9	1.3	984	2	US-08-673-789-9	Sequence 9, Appli
13	9	1.3	984	2	US-08-449-645A-19	Sequence 19, Appl
14	9	1.3	984	2	US-08-702-367A-19	Sequence 19, Appl
15	9	1.3	984	5	PCT-US95-04681-19	Sequence 19, Appl
16	8	1.1	116	1	US-08-434-705B-8	Sequence 8, Appli
17	8	1.1	116	2	US-09-086-201-8	Sequence 8, Appli
18	8	1.1	396	4	US-08-860-656B-2	Sequence 2, Appli
19	8	1.1	396	4	US-08-860-656B-3	Sequence 3, Appli
20	8	1.1	637	2	US-08-426-125-10	Sequence 10, Appl
21	8	1.1	637	2	US-08-455-355-10	Sequence 10, Appl
22	8	1.1	1170	4	US-09-749-588-2	Sequence 2, Appli
23	8	1.1	1209	4	US-09-749-588-4	Sequence 4, Appli
24	7	1.0	13	4	US-08-602-999A-89	Sequence 89, Appl
25	7	1.0	13	4	US-08-278-865-89	Sequence 89, Appl
26	7	1.0	13	4	US-09-500-124-89	Sequence 89, Appl
27	7	1.0	20	4	US-08-860-656B-10	Sequence 10, Appl
28	7	1.0	31	4	US-08-602-999A-57	Sequence 57, Appl
29	7	1.0	31	4	US-08-278-865-57	Sequence 57, Appl
30	7	1.0	31	4	US-09-500-124-57	Sequence 57, Appl
c 31	7	1.0	39	4	US-09-172-841-57	Sequence 57, Appl
c 32	7	1.0	43	3	US-08-899-578-5	Sequence 5, Appli
c 33	7	1.0	45	2	US-08-766-858A-38	Sequence 38, Appl
c 34	7	1.0	51	4	US-09-150-460B-21	Sequence 21, Appl
35	7	1.0	53	4	US-08-905-223-496	Sequence 496, App
c 36	7	1.0	98	1	US-08-308-086-4	Sequence 4, Appli
c 37	7	1.0	98	2	US-08-479-078-5	Sequence 5, Appli
c 38	7	1.0	98	4	US-08-975-040-22	Sequence 22, Appl
c 39	7	1.0	99	1	US-08-202-389-38	Sequence 38, Appl
c 40	7	1.0	101	2	US-08-574-959A-5	Sequence 5, Appli
c 41	7	1.0	101	4	US-09-357-014-5	Sequence 5, Appli
c 42	7	1.0	102	2	US-08-820-754-24	Sequence 24, Appl
c 43	7	1.0	102	3	US-08-956-652-24	Sequence 24, Appl
c 44	7	1.0	102	3	US-08-956-869-24	Sequence 24, Appl
c 45	7	1.0	102	3	US-08-948-547-24	Sequence 24, Appl

ALIGNMENTS

```
RESULT 1
US-08-523-373-22
; Sequence 22, Application US/08523373
; Patent No. 6037145
;  GENERAL INFORMATION:
;    APPLICANT:  Yabuta, Masayuki
;    APPLICANT:  Ohsuye, Kazuhiro
;    TITLE OF INVENTION:  PROCESS FOR PRODUCTION OF PROTEIN
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  699 Prince Street
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22314-3187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/523,373
;      FILING DATE:  05-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-238595
;      FILING DATE:  07-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-296028
;      FILING DATE:  07-NOV-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meuth, Donna M.
;      REGISTRATION NUMBER:  36,607
;      REFERENCE/DOCKET NUMBER:  001560-251
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  703-836-6620
;      TELEFAX:  703-836-2021
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  213 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-523-373-22

Alignment Scores:
Pred. No.:               0.351          Length:        213
Score:                   10.00          Matches:       10
Percent Similarity:      100.00%        Conservative:  0
Best Local Similarity:   100.00%        Mismatches:    0
Query Match:             1.40%          Indels:        0
DB:                      3              Gaps:          0

US-10-008-355-1 (1-2139) x US-08-523-373-22 (1-213)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    165 ThrGlyGlyAsnSerGlySerProValPhe 174

RESULT 2
US-08-523-373-23
; Sequence 23, Application US/08523373
; Patent No. 6037145
;  GENERAL INFORMATION:
;    APPLICANT:  Yabuta, Masayuki
;    APPLICANT:  Ohsuye, Kazuhiro
;    TITLE OF INVENTION:  PROCESS FOR PRODUCTION OF PROTEIN
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  699 Prince Street
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22314-3187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/523,373
;      FILING DATE:  05-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-238595
;      FILING DATE:  07-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-296028
;      FILING DATE:  07-NOV-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meuth, Donna M.
;      REGISTRATION NUMBER:  36,607
;      REFERENCE/DOCKET NUMBER:  001560-251
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6620
;      TELEFAX:  703-836-2021
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  214 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-523-373-23

Alignment Scores:
Pred. No.:               0.351          Length:        214
Score:                   10.00          Matches:       10
Percent Similarity:      100.00%        Conservative:  0
Best Local Similarity:   100.00%        Mismatches:    0
Query Match:             1.40%          Indels:        0
DB:                      3              Gaps:          0

US-10-008-355-1 (1-2139) x US-08-523-373-23 (1-214)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    165 ThrGlyGlyAsnSerGlySerProValPhe 174

RESULT 3
US-08-523-373-24
; Sequence 24, Application US/08523373
; Patent No. 6037145
;  GENERAL INFORMATION:
;    APPLICANT:  Yabuta, Masayuki
;    APPLICANT:  Ohsuye, Kazuhiro
;    TITLE OF INVENTION:  PROCESS FOR PRODUCTION OF PROTEIN
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  699 Prince Street
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22314-3187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/523,373
;      FILING DATE:  05-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-238595
;      FILING DATE:  07-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-296028
;      FILING DATE:  07-NOV-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meuth, Donna M.
;      REGISTRATION NUMBER:  36,607
;      REFERENCE/DOCKET NUMBER:  001560-251
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6620
;      TELEFAX:  703-836-2021
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  215 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-523-373-24

Alignment Scores:
Pred. No.:               0.351          Length:        215
Score:                   10.00          Matches:       10
Percent Similarity:      100.00%        Conservative:  0
Best Local Similarity:   100.00%        Mismatches:    0
Query Match:             1.40%          Indels:        0
DB:                      3              Gaps:          0

US-10-008-355-1 (1-2139) x US-08-523-373-24 (1-215)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    165 ThrGlyGlyAsnSerGlySerProValPhe 174
```

```
RESULT 4
US-08-657-192-3
; Sequence 3, Application US/08657192
; Patent No. 5747321
;  GENERAL INFORMATION:
;    APPLICANT:  YABUTA, Masayuki
;    APPLICANT:  OHSUYE, Kazuhiro
;    TITLE OF INVENTION:  MUTANT STAPHYLOCOCCUS AUREUS V8
;    TITLE OF INVENTION:  PROTEASES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/657,192
;      FILING DATE:  03-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 7-170086
;      FILING DATE:  02-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meuth, Donna M.
;      REGISTRATION NUMBER:  36,607
;      REFERENCE/DOCKET NUMBER:  001560-264
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  344 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-657-192-3

Alignment Scores:
Pred. No.:               0.332         Length:        344
Score:                   10.00         Matches:       10
Percent Similarity:      100.00%       Conservative:  0
Best Local Similarity:   100.00%       Mismatches:    0
Query Match:             1.40%         Indels:        0
DB:                      1             Gaps:          0

US-10-008-355-1 (1-2139) x US-08-657-192-3 (1-344)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 5
US-08-523-373-5
; Sequence 5, Application US/08523373
; Patent No. 6037145
;  GENERAL INFORMATION:
;    APPLICANT:  Yabuta, Masayuki
;    APPLICANT:  Ohsuye, Kazuhiro
;    TITLE OF INVENTION:  PROCESS FOR PRODUCTION OF PROTEIN
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  699 Prince Street
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22314-3187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/523,373
;      FILING DATE:  05-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-238595
;      FILING DATE:  07-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-296028
;      FILING DATE:  07-NOV-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meuth, Donna M.
;      REGISTRATION NUMBER:  36,607
;      REFERENCE/DOCKET NUMBER:  001560-251
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6620
;      TELEFAX:  703-836-2021
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  344 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-523-373-5

Alignment Scores:
Pred. No.:               0.332         Length:        344
Score:                   10.00         Matches:       10
Percent Similarity:      100.00%       Conservative:  0
Best Local Similarity:   100.00%       Mismatches:    0
Query Match:             1.40%         Indels:        0
DB:                      3             Gaps:          0

US-10-008-355-1 (1-2139) x US-08-523-373-5 (1-344)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 6
US-08-523-373-6
; Sequence 6, Application US/08523373
; Patent No. 6037145
;  GENERAL INFORMATION:
;    APPLICANT:  Yabuta, Masayuki
;    APPLICANT:  Ohsuye, Kazuhiro
;    TITLE OF INVENTION:  PROCESS FOR PRODUCTION OF PROTEIN
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  699 Prince Street
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22314-3187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/523,373
;      FILING DATE:  05-SEP-1995
;      CLASSIFICATION:  435
```

5/5/98

30

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-238595
;      FILING DATE:  07-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-296028
;      FILING DATE:  07-NOV-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meuth, Donna M.
;      REGISTRATION NUMBER:  36,607
;      REFERENCE/DOCKET NUMBER:  001560-251
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6620
;      TELEFAX:  703-836-2021
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  392 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-523-373-6

Alignment Scores:
Pred. No.:              0.327         Length:        392
Score:                  10.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.40%         Indels:        0
DB:                     3             Gaps:          0

US-10-008-355-1 (1-2139) x US-08-523-373-6 (1-392)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 7
US-08-657-192-9
; Sequence 9, Application US/08657192
; Patent No. 5747321
;  GENERAL INFORMATION:
;    APPLICANT:  YABUTA, Masayuki
;    APPLICANT:  OHSUYE, Kazuhiro
;    TITLE OF INVENTION:  MUTANT STAPHYLOCOCCUS AUREUS V8
;    TITLE OF INVENTION:  PROTEASES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/657,192
;      FILING DATE:  03-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 7-170086,
;      FILING DATE:  02-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meuth, Donna M.
;      REGISTRATION NUMBER:  36,607
;      REFERENCE/DOCKET NUMBER:  001560-264
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  532 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-657-192-9

Alignment Scores:
Pred. No.:              0.315         Length:        532
Score:                  10.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.40%         Indels:        0
DB:                     1             Gaps:          0

US-10-008-355-1 (1-2139) x US-08-657-192-9 (1-532)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 8
US-08-523-373-7
; Sequence 7, Application US/08523373
; Patent No. 6037145
;  GENERAL INFORMATION:
;    APPLICANT:  Yabuta, Masayuki
;    APPLICANT:  Ohsuye, Kazuhiro
;    TITLE OF INVENTION:  PROCESS FOR PRODUCTION OF PROTEIN
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  699 Prince Street
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22314-3187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/523,373
;      FILING DATE:  05-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-238595
;      FILING DATE:  07-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-296028
;      FILING DATE:  07-NOV-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meuth, Donna M.
;      REGISTRATION NUMBER:  36,607
;      REFERENCE/DOCKET NUMBER:  001560-251
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-836-6620
;      TELEFAX:  703-836-2021
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  532 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-523-373-7

Alignment Scores:
Pred. No.:              0.315         Length:        532
Score:                  10.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  0
```

```
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.40%          Indels:        0
DB:                     3              Gaps:          0

US-10-008-355-1 (1-2139) x US-08-523-373-7 (1-532)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db     289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 9
US-08-657-192-15
; Sequence 15, Application US/08657192
; Patent No. 5747321
;  GENERAL INFORMATION:
;    APPLICANT:  YABUTA, Masayuki
;    APPLICANT:  OHSUYE, Kazuhiro
;    TITLE OF INVENTION:  MUTANT STAPHYLOCOCCUS AUREUS V8
;    TITLE OF INVENTION:  PROTEASES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/657,192
;      FILING DATE:  03-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 7-170086
;      FILING DATE:  02-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meuth, Donna M.
;      REGISTRATION NUMBER:  36,607
;      REFERENCE/DOCKET NUMBER:  001560-264
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  537 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-657-192-15

Alignment Scores:
Pred. No.:              0.314          Length:        537
Score:                  10.00          Matches:       10
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.40%          Indels:        0
DB:                     1              Gaps:          0

US-10-008-355-1 (1-2139) x US-08-657-192-15 (1-537)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db     289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 10
US-08-332-562A-67
; Sequence 67, Application US/08332562A
; Patent No. 5985599
;  GENERAL INFORMATION:
;    APPLICANT:  McKENZIE, Ian F.C.
;    APPLICANT:  HOGARTH, Mark P.
;    APPLICANT:  HIBBS, Margaret L.
;    APPLICANT:  SCOTT, Bernadette M.
;    APPLICANT:  BONADONNA, Lisa
;    APPLICANT:  HULETT, Mark D.
;    TITLE OF INVENTION:  FC RECEPTOR FOR IMMUNOGLOBULIN.
;    NUMBER OF SEQUENCES:  136
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,562A
;      FILING DATE:  31-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/896,457
;      FILING DATE:  27-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  54270/119/GRHA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  67:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-332-562A-67

Alignment Scores:
Pred. No.:              4.44           Length:        15
Score:                  9.00           Matches:       9
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.26%          Indels:        0
DB:                     2              Gaps:          0

US-10-008-355-1 (1-2139) x US-08-332-562A-67 (1-15)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTA 1956
           |||||||||||||||||||||||||||
Db       5 ThrGlyGlyAsnSerGlySerProVal 13

RESULT 11
US-09-134-001C-3729
; Sequence 3729, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
```

```
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3729
;   LENGTH: 311
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3729

Alignment Scores:
Pred. No.:               3.08           Length:        311
Score:                   9.00           Matches:       9
Percent Similarity:      100.00%        Conservative:  0
Best Local Similarity:   100.00%        Mismatches:    0
Query Match:             1.26%          Indels:        0
DB:                      4              Gaps:          0

US-10-008-355-1 (1-2139) x US-09-134-001C-3729 (1-311)

Qy   1933 GGCGGTAACTCCGGTAGCCCCGTATTC 1959
          |||||||||||||||||||||||||||
Db    261 GlyGlyAsnSerGlySerProValPhe 269

RESULT 12
US-08-673-789-9
; Sequence 9, Application US/08673789
; Patent No. 5814479
;  GENERAL INFORMATION:
;    APPLICANT:  ZHOU, RENPING; SCHULZ, NICHOLAS,
;    APPLICANT:  T.; KROMER, LAWRENCE, F.; VANDE WOUDE,
;    APPLICANT:  GEORGE, F.
;    TITLE OF INVENTION:  BSK RECEPTOR LIKE
;    TITLE OF INVENTION:  TYROSINE KINASE AND LIGAND AND THEIR
;    TITLE OF INVENTION:  USE IN DIAGNOSTIC AND THERAPEUTIC
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/673,789
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/177,812
;      FILING DATE:  04-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  37,341
;      REFERENCE/DOCKET NUMBER:  2026-4105
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  984
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  UNKNOWN
;      TOPOLOGY:  UNKNOWN
US-08-673-789-9

Alignment Scores:
Pred. No.:               2.68           Length:        984
Score:                   9.00           Matches:       9
Percent Similarity:      100.00%        Conservative:  0
Best Local Similarity:   100.00%        Mismatches:    0
Query Match:             1.26%          Indels:        0
DB:                      2              Gaps:          0

US-10-008-355-1 (1-2139) x US-08-673-789-9 (1-984)

Qy     25 CTTCTCGGAGCAGCCCTGCTGTTGGGT 51
          |||||||||||||||||||||||||||
Db    556 LeuLeuGlyAlaAlaLeuLeuLeuGly 564

RESULT 13
US-08-449-645A-19
; Sequence 19, Application US/08449645A
; Patent No. 5981245
;  GENERAL INFORMATION:
;    APPLICANT:  Fox, Gary M.
;    TITLE OF INVENTION:  EPH-Like Receptor Protein Tyrosine
;    TITLE OF INVENTION:  Kinases
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amgen Patent Operations/RBW
;      STREET:  1840 Dehavilland Drive
;      CITY:  Thousand Oaks
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/449,645A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Winter, Robert B.
;      REFERENCE/DOCKET NUMBER:  A-287
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  984 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-449-645A-19

Alignment Scores:
Pred. No.:               2.68           Length:        984
Score:                   9.00           Matches:       9
Percent Similarity:      100.00%        Conservative:  0
Best Local Similarity:   100.00%        Mismatches:    0
Query Match:             1.26%          Indels:        0
DB:                      2              Gaps:          0

US-10-008-355-1 (1-2139) x US-08-449-645A-19 (1-984)

Qy     25 CTTCTCGGAGCAGCCCTGCTGTTGGGT 51
          |||||||||||||||||||||||||||
Db    556 LeuLeuGlyAlaAlaLeuLeuLeuGly 564

RESULT 14
US-08-702-367A-19
; Sequence 19, Application US/08702367A
; Patent No. 5981246
;  GENERAL INFORMATION:
;    APPLICANT:  Fox, Gary M.
;    TITLE OF INVENTION:  EPH-Like Receptor Protein Tyrosine
;    TITLE OF INVENTION:  Kinases
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amgen Patent Operations/RBW
;      STREET:  1840 Dehavilland Drive
```

```
;      CITY:  Thousand Oaks
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/702,367A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Winter, Robert B.
;      REFERENCE/DOCKET NUMBER:  A-287
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  984 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-702-367A-19

Alignment Scores:
Pred. No.:              2.68          Length:        984
Score:                  9.00          Matches:       9
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.26%         Indels:        0
DB:                     2             Gaps:          0

US-10-008-355-1 (1-2139) x US-08-702-367A-19 (1-984)

Qy      25 CTTCTCGGAGCAGCCCTGCTGTTGGGT 51
           |||||||||||||||||||||||||||
Db     556 LeuLeuGlyAlaAlaLeuLeuLeuGly 564

RESULT 15
PCT-US95-04681-19
; Sequence 19, Application PC/TUS9504681
;  GENERAL INFORMATION:
;    APPLICANT:  Fox, Gary M.
;    TITLE OF INVENTION:  EPH-Like Receptor Protein Tyrosine
;    TITLE OF INVENTION:  Kinases
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amgen Patent Operations/RBW
;      STREET:  1840 Dehavilland Drive
;      CITY:  Thousand Oaks
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91320
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04681
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Winter, Robert B.
;      REFERENCE/DOCKET NUMBER:  A-287
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  984 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-04681-19

Alignment Scores:
Pred. No.:              2.68          Length:        984
Score:                  9.00          Matches:       9
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.26%         Indels:        0
DB:                     5             Gaps:          0

US-10-008-355-1 (1-2139) x PCT-US95-04681-19 (1-984)

Qy      25 CTTCTCGGAGCAGCCCTGCTGTTGGGT 51
           |||||||||||||||||||||||||||
Db     556 LeuLeuGlyAlaAlaLeuLeuLeuGly 564

Search completed: May 16, 2003, 13:17:28
Job time : 44 secs
```

DP

```
                     GenCore version 5.1.4_p5_4578
            Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:          May 16, 2003, 12:38:22 ; Search time 113.5 Seconds
                                          (without alignments)
                                          5022.433 Million cell updates/sec

Title:           US-10-008-355-1
Perfect score:   712
Sequence:        1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:   OLIGO
                 Xgapop 60.0 , Xgapext 60.0
                 Ygapop 60.0 , Ygapext 60.0
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        908470 seqs, 133250620 residues

Word size:       1

Total number of hits satisfying chosen parameters:      1687582

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110425_9716/app_query.fasta_1.2311
-DB=A_Geneseq_101002 -QFMT=fastan -SUFFIX=olin2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_153_@runat_16052003_110425_9716 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :       A_Geneseq_101002:*
                 1:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
                 2:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
                 3:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
                 4:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
                 5:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
                 6:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
                 7:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
                 8:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
                 9:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
                 10:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
                 11:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
                 12:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
                 13:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
                 14:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
                 15:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
                 16:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
                 17:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
                 18:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
                 19:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
                 20:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
                 21:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
                 22:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
                 23:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	712	100.0	712	23	AAO15205	Porphyromonas ging
	2	52	7.3	52	23	AAO15206	Porphyromonas ging
	3	21	2.9	26	23	AAO15221	Porphyromonas ging
	4	10	1.4	10	23	AAO15228	Porphyromonas ging
	5	10	1.4	52	23	AAO15207	Staphylococcus aur
	6	10	1.4	213	17	AAR91042	V8 mature protease
	7	10	1.4	214	17	AAR91043	V8 mature protease
	8	10	1.4	215	17	AAR91044	V8 mature protease
	9	10	1.4	336	13	AAR29644	Protease from S. A
	10	10	1.4	344	17	AAR91033	Beta-galactosidase
	11	10	1.4	344	18	AAW22218	Protein encoded by
	12	10	1.4	357	13	AAR26842	Protease from S. A
	13	10	1.4	392	17	AAR91034	Beta-galactosidase
	14	10	1.4	532	17	AAR91035	Recombinant V8 pro
	15	10	1.4	532	18	AAW22219	Protein encoded by
	16	10	1.4	537	18	AAW22220	Protein encoded by
	17	9	1.3	9	23	AAO15229	Porphyromonas ging
	18	9	1.3	76	22	AAU63245	Propionibacterium
	19	9	1.3	282	22	AAG81635	S. epidermidis ope
	20	9	1.3	311	23	ABP38884	Staphylococcus epi
	21	8	1.1	27	23	ABB96842	Omega-conopeptide
	22	8	1.1	55	21	AAB35780	Human CART peptide
	23	8	1.1	73	23	ABB96626	Omega-conopeptide
	24	8	1.1	75	21	AAB35779	Human CART peptide
c	25	8	1.1	80	22	AAU50283	Propionibacterium
	26	8	1.1	88	21	AAB35772	Human truncated CA
c	27	8	1.1	89	22	AAU66041	Propionibacterium
	28	8	1.1	116	17	AAW06814	Human cocaine and
	29	8	1.1	116	19	AAW78906	Human cocaine and
	30	8	1.1	116	20	AAY45180	Human cocaine and
	31	8	1.1	116	20	AAW81341	Human cocaine and
c	32	8	1.1	138	22	AAO07965	Human polypeptide
c	33	8	1.1	152	22	AAB94914	Human protein sequ
c	34	8	1.1	159	22	AAM92139	Human digestive sy
	35	8	1.1	173	22	ABG18197	Novel human diagno
c	36	8	1.1	211	22	ABB67468	Drosophila melanog
	37	8	1.1	236	23	AAU93181	Arabidopsis transc
	38	8	1.1	306	22	ABG27777	Novel human diagno
c	39	8	1.1	312	22	AAB87739	Human T2R09 amino
	40	8	1.1	335	23	ABP30343	Streptococcus poly
	41	8	1.1	340	23	ABP28097	Streptococcus poly
	42	8	1.1	340	23	ABP29857	Streptococcus poly
c	43	8	1.1	395	22	ABB52640	Escherichia coli p
	44	8	1.1	396	17	AAR97569	Interleukin-2 rece
	45	8	1.1	413	22	ABB11836	Human IL-2R-associ

ALIGNMENTS

```
RESULT 1
AAO15205
ID   AAO15205 standard; Protein; 712 AA.
XX
AC   AAO15205;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7).
XX
KW   Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
KW   DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW   periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200238742-A2.
XX
PD   16-MAY-2002.
```

```
XX
PF   08-NOV-2001; 2001WO-US46782.
XX
PR   08-NOV-2000; 2000US-246827P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,  Potempa JS,  Banbula A,  Bugno M;
XX
DR   WPI; 2002-490075/52.
DR   N-PSDB; AAL43635.
XX
PT   Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT   the dipeptidylpeptidase for protecting an animal from periodontal
PT   disease caused by Porphyromonas gingivalis  -
XX
PS   Claim 7; Fig 4; 65pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC   enzyme has amidolytic activity for cleavage of a peptide bond between the
CC   second and third amino acids from the N-terminal end of a target peptide.
CC   The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC   substituent on the alpha-carbon atom of the second amino acid from the
CC   N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC   useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC   reducing the growth of a bacterium and protecting an animal from a
CC   periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC   or periodontitis). The present amino acid sequence represents the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the
CC   invention.
XX
SQ   Sequence   712 AA;

Alignment Scores:
Pred. No.:               0              Length:        712
Score:                   712.00         Matches:       712
Percent Similarity:      100.00%        Conservative:  0
Best Local Similarity:   100.00%        Mismatches:    0
Query Match:             100.00%        Indels:        0
DB:                      23             Gaps:          0

US-10-008-355-1 (1-2139) x AA015205 (1-712)

Qy       1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20

Qy      61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAsp 40

Qy     121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60

Qy     181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 SerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80

Qy     241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100

Qy     301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuPro 120

Qy     361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140

Qy     421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160

Qy     481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180

Qy     541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200

Qy     601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro 220

Qy     661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240

Qy     721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln 260

Qy     781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280

Qy     841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300

Qy     901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320

Qy     961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340

Qy    1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360

Qy    1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380

Qy    1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400

Qy    1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420

Qy    1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSer 440

Qy    1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAsp 460

Qy    1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480

Qy    1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500

Qy    1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCACTAGAG 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520

Qy   1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540

Qy   1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560

Qy   1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580

Qy   1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600

Qy   1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620

Qy   1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640

Qy   1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660

Qy   1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680

Qy   2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGly 700

Qy   2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATC 2136
          ||||||||||||||||||||||||||||||||||||
Db    701 GlnCysProArgLeuIleGlnGluLeuLysLeuIle 712

RESULT 2
AAO15206
ID   AAO15206 standard; Protein; 52 AA.
XX
AC   AAO15206;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) C-terminal region.
XX
KW   Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
KW   DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW   periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200238742-A2.
XX
PD   16-MAY-2002.
XX
PF   08-NOV-2001; 2001WO-US46782.
XX
PR   08-NOV-2000; 2000US-246827P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,  Potempa JS,  Banbula A,  Bugno M;
XX
DR   WPI; 2002-490075/52.
XX
PT   Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT   the dipeptidylpeptidase for protecting an animal from periodontal
PT   disease caused by Porphyromonas gingivalis  -
XX
PS   Example 6; Fig 5; 65pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC   enzyme has amidolytic activity for cleavage of a peptide bond between the
CC   second and third amino acids from the N-terminal end of a target peptide.
CC   The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC   substituent on the alpha-carbon atom of the second amino acid from the
CC   N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC   useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC   reducing the growth of a bacterium and protecting an animal from a
CC   periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC   or periodontitis). The present amino acid sequence represents the
CC   C-terminal region of the Porphyromonas gingivalis dipeptidylpeptidase-7
CC   (DPP-7) enzyme.
XX
SQ   Sequence   52 AA;

Alignment Scores:
Pred. No.:              1.17e-44      Length:        52
Score:                  52.00         Matches:       52
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            7.30%         Indels:        0
DB:                     23            Gaps:          0

US-10-008-355-1 (1-2139) x AAO15206 (1-52)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCT 1989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAla 20

Qy   1990 TTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGC 2049
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArg 40

Qy   2050 ACAATCAGCGTGGACATCCGCTACGTTCTCTTCATG 2085
          ||||||||||||||||||||||||||||||||||||
Db     41 ThrIleSerValAspIleArgTyrValLeuPheMet 52

RESULT 3
AAO15221
ID   AAO15221 standard; Peptide; 26 AA.
XX
AC   AAO15221;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) N-terminal region.
XX
KW   Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
KW   DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW   periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200238742-A2.
XX
PD   16-MAY-2002.
XX
PF   08-NOV-2001; 2001WO-US46782.
XX
PR   08-NOV-2000; 2000US-246827P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,  Potempa JS,  Banbula A,  Bugno M;
XX
DR   WPI; 2002-490075/52.
XX
```

```
PT   Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT   the dipeptidylpeptidase for protecting an animal from periodontal
PT   disease caused by Porphyromonas gingivalis  -
XX
PS   Example 6; Page 30; 65pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC   enzyme has amidolytic activity for cleavage of a peptide bond between the
CC   second and third amino acids from the N-terminal end of a target peptide.
CC   The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC   substituent on the alpha-carbon atom of the second amino acid from the
CC   N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC   useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC   reducing the growth of a bacterium and protecting an animal from a
CC   periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC   or periodontitis). The present amino acid sequence represents the
CC   N-terminal region of the Porphyromonas gingivalis dipeptidylpeptidase-7
CC   (DPP-7) enzyme.
XX
SQ   Sequence   26 AA;

Alignment Scores:
Pred. No.:              2.33e-12       Length:         26
Score:                  21.00          Matches:        21
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            2.95%          Indels:         0
DB:                     23             Gaps:           0

US-10-008-355-1 (1-2139) x AAO15221 (1-26)

Qy      82 ATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTT 141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6 MetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPhe 25

Qy     142 ACG 144
           |||
Db      26 Thr 26

RESULT 4
AAO15228
ID   AAO15228 standard; Peptide; 10 AA.
XX
AC   AAO15228;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) active site 1.
XX
KW   Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage; active site;
KW   DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW   periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200238742-A2.
XX
PD   16-MAY-2002.
XX
PF   08-NOV-2001; 2001WO-US46782.
XX
PR   08-NOV-2000; 2000US-246827P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,  Potempa JS,  Banbula A,  Bugno M;
XX
DR   WPI; 2002-490075/52.
XX
PT   Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT   the dipeptidylpeptidase for protecting an animal from periodontal
PT   disease caused by Porphyromonas gingivalis  -
XX
PS   Claim 5; Page 32; 65pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC   enzyme has amidolytic activity for cleavage of a peptide bond between the
CC   second and third amino acids from the N-terminal end of a target peptide.
CC   The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC   substituent on the alpha-carbon atom of the second amino acid from the
CC   N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC   useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC   reducing the growth of a bacterium and protecting an animal from a
CC   periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC   or periodontitis). The present amino acid sequence represents an active
CC   site region of the Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7)
CC   enzyme.
XX
SQ   Sequence   10 AA;

Alignment Scores:
Pred. No.:              0.72           Length:         10
Score:                  10.00          Matches:        10
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            1.40%          Indels:         0
DB:                     23             Gaps:           0

US-10-008-355-1 (1-2139) x AAO15228 (1-10)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db       1 ThrGlyGlyAsnSerGlySerProValPhe 10

RESULT 5
AAO15207
ID   AAO15207 standard; Protein; 52 AA.
XX
AC   AAO15207;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Staphylococcus aureus V8 endopeptidase C-terminal region.
XX
KW   Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
KW   DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW   periodontitis; V8 endopeptidease.
XX
OS   Staphylococcus aureus.
XX
PN   WO200238742-A2.
XX
PD   16-MAY-2002.
XX
PF   08-NOV-2001; 2001WO-US46782.
XX
PR   08-NOV-2000; 2000US-246827P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,  Potempa JS,  Banbula A,  Bugno M;
XX
DR   WPI; 2002-490075/52.
XX
PT   Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT   the dipeptidylpeptidase for protecting an animal from periodontal
PT   disease caused by Porphyromonas gingivalis  -
XX
PS   Example 6; Fig 5; 65pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC   enzyme has amidolytic activity for cleavage of a peptide bond between the
CC   second and third amino acids from the N-terminal end of a target peptide.
```

```
CC   The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC   substituent on the alpha-carbon atom of the second amino acid from the
CC   N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC   useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC   reducing the growth of a bacterium and protecting an animal from a
CC   periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC   or periodontitis). The present amino acid sequence represents the
CC   C-terminal region of the Staphylococcus aureus V8 endopeptidase enzyme.
XX
SQ   Sequence   52 AA;

Alignment Scores:
Pred. No.:              0.618         Length:         52
Score:                  10.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            1.40%         Indels:         0
DB:                     23            Gaps:           0

US-10-008-355-1 (1-2139) x AAO15207 (1-52)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db       1 ThrGlyGlyAsnSerGlySerProValPhe 10

RESULT 6
AAR91042
ID   AAR91042 standard; Peptide; 213 AA.
XX
AC   AAR91042;
XX
DT   23-MAY-1996  (first entry)
XX
DE   V8 mature protease (aa1-213).
XX
KW   Linker peptide; V8 protease; Staphylococcus aureus;
KW   recombinant protein; fusion protein; beta-galactosidase;
KW   Escherichia coli; transposon Tn903;
KW   aminoglycoside 3'-phosphotransferase.
XX
OS   Staphylococcus aureus strain V8 (ATCC 27733).
XX
PN   EP700995-A2.
XX
PD   13-MAR-1996.
XX
PF   06-SEP-1995;   95EP-0306235.
XX
PR   07-NOV-1994;   94JP-0296028.
PR   07-SEP-1994;   94JP-0238595.
XX
PA   (SUNR ) SUNTORY LTD.
XX
PI   Ohsuye K,  Yabuta M;
XX
DR   WPI; 1996-141021/15.
XX
PT   Prodn. of recombinant polypeptide(s) - using host cells transformed
PT   with a gene coding for the desired polypeptide fused to a protective
PT   polypeptide
XX
PS   Example 8; Fig 14a; 44pp; English.
XX
CC   A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
CC   and C are protective polypeptides, B is Staphylococcus aureus mature
CC   V8 protease lacking the C-terminal repeat region, and L is a linker
CC   (AAR91032).  It was produced as an inclusion body in Escherichia coli
CC   host cells, and was cleaved using the E. coli OmpT protease to
CC   yield active V8 protease.  Extension of the C-terminal end of the
CC   V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)
CC   i.e. up to residue Phe-215, also resulted in the formation of
CC   inclusion bodies in E. coli.  Any further extension gave a soluble
CC   product which exhibited protease activity that repressed growth of
CC   the host cells.
XX
SQ   Sequence   213 AA;

Alignment Scores:
Pred. No.:              0.542         Length:         213
Score:                  10.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            1.40%         Indels:         0
DB:                     17            Gaps:           0

US-10-008-355-1 (1-2139) x AAR91042 (1-213)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db     165 ThrGlyGlyAsnSerGlySerProValPhe 174

RESULT 7
AAR91043
ID   AAR91043 standard; Peptide; 214 AA.
XX
AC   AAR91043;
XX
DT   23-MAY-1996  (first entry)
XX
DE   V8 mature protease (aa1-214).
XX
KW   Linker peptide; V8 protease; Staphylococcus aureus;
KW   recombinant protein; fusion protein; beta-galactosidase;
KW   Escherichia coli; transposon Tn903;
KW   aminoglycoside 3'-phosphotransferase.
XX
OS   Staphylococcus aureus strain V8 (ATCC 27733).
XX
PN   EP700995-A2.
XX
PD   13-MAR-1996.
XX
PF   06-SEP-1995;   95EP-0306235.
XX
PR   07-NOV-1994;   94JP-0296028.
PR   07-SEP-1994;   94JP-0238595.
XX
PA   (SUNR ) SUNTORY LTD.
XX
PI   Ohsuye K,  Yabuta M;
XX
DR   WPI; 1996-141021/15.
XX
PT   Prodn. of recombinant polypeptide(s) - using host cells transformed
PT   with a gene coding for the desired polypeptide fused to a protective
PT   polypeptide
XX
PS   Example 8; Fig 14b; 44pp; English.
XX
CC   A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
CC   and C are protective polypeptides, B is Staphylococcus aureus mature
CC   V8 protease lacking the C-terminal repeat region, and L is a linker
CC   (AAR91032).  It was produced as an inclusion body in Escherichia coli
CC   host cells, and was cleaved using the E. coli OmpT protease to
CC   yield active V8 protease.  Extension of the C-terminal end of the
CC   V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)
CC   i.e. up to residue Phe-215, also resulted in the formation of
CC   inclusion bodies in E. coli.  Any further extension gave a soluble
CC   product which exhibited protease activity that repressed growth of
CC   the host cells.
XX
SQ   Sequence   214 AA;

Alignment Scores:
Pred. No.:              0.541         Length:         214
Score:                  10.00         Matches:        10
```

29 '96

```
Percent Similarity:      100.00%       Conservative:   0
Best Local Similarity:   100.00%       Mismatches:     0
Query Match:             1.40%         Indels:         0
DB:                      17            Gaps:           0

US-10-008-355-1 (1-2139) x AAR91043 (1-214)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db     165 ThrGlyGlyAsnSerGlySerProValPhe 174

RESULT 8
AAR91044
ID   AAR91044 standard; Peptide; 215 AA.
XX
AC   AAR91044;
XX
DT   23-MAY-1996  (first entry)
XX
DE   V8 mature protease (aa1-215).
XX
KW   Linker peptide; V8 protease; Staphylococcus aureus;
KW   recombinant protein; fusion protein; beta-galactosidase;
KW   Escherichia coli; transposon Tn903;
KW   aminoglycoside 3'-phosphotransferase.
XX
OS   Staphylococcus aureus strain V8 (ATCC 27733).
XX
PN   EP700995-A2.
XX
PD   13-MAR-1996.
XX
PF   06-SEP-1995;   95EP-0306235.
XX
PR   07-NOV-1994;   94JP-0296028.
PR   07-SEP-1994;   94JP-0238595.
XX
PA   (SUNR ) SUNTORY LTD.
XX
PI   Ohsuye K,  Yabuta M;
XX
DR   WPI; 1996-141021/15.
XX
PT   Prodn. of recombinant polypeptide(s) - using host cells transformed
PT   with a gene coding for the desired polypeptide fused to a protective
PT   polypeptide
XX
PS   Example 8; Fig 14c; 44pp; English.
XX
CC   A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
CC   and C are protective polypeptides, B is Staphylococcus aureus mature
CC   V8 protease lacking the C-terminal repeat region, and L is a linker
CC   (AAR91032).  It was produced as an inclusion body in Escherichia coli
CC   host cells, and was cleaved using the E. coli OmpT protease to
CC   yield active V8 protease.  Extension of the C-terminal end of the
CC   V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)
CC   i.e. up to residue Phe-215, also resulted in the formation of
CC   inclusion bodies in E. coli.  Any further extension gave a soluble
CC   product which exhibited protease activity that repressed growth of
CC   the host cells.
XX
SQ   Sequence   215 AA;

Alignment Scores:
Pred. No.:               0.541         Length:         215
Score:                   10.00         Matches:        10
Percent Similarity:      100.00%       Conservative:   0
Best Local Similarity:   100.00%       Mismatches:     0
Query Match:             1.40%         Indels:         0
DB:                      17            Gaps:           0

US-10-008-355-1 (1-2139) x AAR91044 (1-215)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db     165 ThrGlyGlyAsnSerGlySerProValPhe 174

RESULT 9
AAR29644
ID   AAR29644 standard; Protein; 336 AA.
XX
AC   AAR29644;
XX
DT   11-FEB-1993  (first entry)
XX
DE   Protease from S. Aureus.
XX
KW   Protease; PCR; amplify; Staphylococcus.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..68
FT                   /label= Signal_peptide
FT   Protein         69..336
FT                   /label= Protease
XX
PN   JP04211370-A.
XX
PD   03-AUG-1992.
XX
PF   19-FEB-1991;   91JP-0024633.
XX
PR   20-FEB-1990;   90JP-0040398.
XX
PA   (SHIO ) SHIONOGI & CO LTD.
XX
DR   WPI; 1992-304938/37.
DR   N-PSDB; AAQ27988.
XX
PT   Novel protease prepd. using Bacillus or Saccharomyces host -
PT   capable of cleaving peptide bond at carboxyl terminus of glutamic
PT   acid residues in polypeptide(s)
XX
PS   Disclosure; Page 15-16; 25pp; Japanese.
XX
CC   The sequences given in AAR26842 and AAR29644 are proteases which were
CC   isolated from Staphylococcus aureus strains.  The DNA sequences
CC   encoding these proteins were isolated by PCR using the primer
CC   sequences given in AAQ27960-86.  The protease specifically cleaves
CC   the peptide bond at the C-terminus of the glutamic acid residue in
CC   polypeptide.
XX
SQ   Sequence   336 AA;

Alignment Scores:
Pred. No.:               0.519         Length:         336
Score:                   10.00         Matches:        10
Percent Similarity:      100.00%       Conservative:   0
Best Local Similarity:   100.00%       Mismatches:     0
Query Match:             1.40%         Indels:         0
DB:                      13            Gaps:           0

US-10-008-355-1 (1-2139) x AAR29644 (1-336)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db     233 ThrGlyGlyAsnSerGlySerProValPhe 242

RESULT 10
AAR91033
ID   AAR91033 standard; Protein; 344 AA.
XX
AC   AAR91033;
XX
DT   23-MAY-1996  (first entry)
```

```
XX
DE   Beta-galactosidase-V8 protease fusion protein.
XX
KW   V8 protease; Staphylococcus aureus; recombinant protein;
KW   fusion protein; beta-galactosidase; Escherichia coli.
XX
OS   Chimeric Escherichia coli;
OS   Chimeric Staphylococcus aureus strain V8 (ATCC 27733).
XX
FH   Key             Location/Qualifiers
FT   Region          1..124
FT                   /note= "beta-galactosidase region"
FT   Region          125..344
FT                   /note= "mature V8 protease without the repeat
FT                          region"
XX
PN   EP700995-A2.
XX
PD   13-MAR-1996.
XX
PF   06-SEP-1995;   95EP-0306235.
XX
PR   07-NOV-1994;   94JP-0296028.
PR   07-SEP-1994;   94JP-0238595.
XX
PA   (SUNR ) SUNTORY LTD.
XX
PI   Ohsuye K,  Yabuta M;
XX
DR   WPI; 1996-141021/15.
XX
PT   Prodn. of recombinant polypeptide(s) - using host cells transformed
PT   with a gene coding for the desired polypeptide fused to a protective
PT   polypeptide
XX
PS   Example 2; Page 12-13; 44pp; English.
XX
CC   Fusion proteins (AAR91033 and AAR91034) were constructed comprising
CC   an Escherichia coli beta-galactosidase derivative (protective
CC   polypeptide) fused at its C-terminal end to the S. aureus mature
CC   V8 protease without or with the repeat region.  The constructs
CC   were inserted into vector pG97S4DhCT(G)R6, yielding pV8RPT(-) and
CC   pV8RPT(+), respectively.  Both constructs yielded active protease
CC   when expressed in E. coli JM101 transformants.
XX
SQ   Sequence   344 AA;

Alignment Scores:
Pred. No.:              0.518         Length:         344
Score:                  10.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            1.40%         Indels:         0
DB:                     17            Gaps:           0

US-10-008-355-1 (1-2139) x AAR91033 (1-344)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 11
AAW22218
ID   AAW22218 standard; Protein; 344 AA.
XX
AC   AAW22218;
XX
DT   11-SEP-1997  (first entry)
XX
DE   Protein encoded by pV8RPT(-) construct.
XX
KW   Mutant; Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;
KW   truncation; wild type; PCR; polymerase chain reaction; amplification;
KW   proteolytic activity; fusion protein; beta-galactosidase; urea.
XX
OS   Chimeric - Escherichia coli.
OS   Chimeric - Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..100
FT                   /note= "E. coli beta-galactosidase portion"
FT   Region          101..120
FT                   /note= "synthetic R6 linker"
FT   Region          125..344
FT                   /note= "truncated S. aureus V8 protease portion"
XX
PN   EP745669-A2.
XX
PD   04-DEC-1996.
XX
PF   31-MAY-1996;   96EP-0303939.
XX
PR   02-JUN-1995;   95JP-0170086.
XX
PA   (SUNR ) SUNTORY LTD.
XX
PI   Ohsuye K,  Yabuta M;
XX
DR   WPI; 1997-013693/02.
XX
PT   Staphylococcus aureus V8 protease mutants - with increased
PT   resistance to denaturation
XX
PS   Claim 2; Page 13-14; 42pp; English.
XX
CC   The invention relates to new mutant Staphylococcus aureus V8 proteases
CC   which have enzyme activity even under environmental conditions which
CC   promote protein denaturation.  The mutants are based on 3 truncated V8
CC   proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino
CC   acids from the C-terminal of the wild type protease.  The mutants also
CC   contain amino acid substitutions, especially D44E, N71S and/or R147K.
CC   The protein sequence shown here represents a chimaeric protein
CC   comprising a truncated Staphylococcus aureus V8 protease lacking the
CC   prepro and C-terminal 48 amino acids linked, via a synthetic linker,
CC   downstream of the E. coli beta-galactosidase.  The S. aureus portion of
CC   the chimaera was amplified by the primers AAT73254-5 from wild type
CC   sequence.  The coding sequence was then used to generate mutants of the
CC   V8 protease which retain their levels of activity in the presence of a
CC   higher concentration of protein denaturant e.g. 5 M urea.
XX
SQ   Sequence   344 AA;

Alignment Scores:
Pred. No.:              0.518         Length:         344
Score:                  10.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            1.40%         Indels:         0
DB:                     18            Gaps:           0

US-10-008-355-1 (1-2139) x AAW22218 (1-344)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 12
AAR26842
ID   AAR26842 standard; Protein; 357 AA.
XX
AC   AAR26842;
XX
DT   11-FEB-1993  (first entry)
XX
DE   Protease from S. Aureus ATCC12600.
XX
```

29 '93

```
KW   Protease; PCR; amplify; Staphylococcus.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..68
FT                   /label= Signal_peptide
FT   Protein         69..358
FT                   /label= Protease
XX
PN   JP04211370-A.
XX
PD   03-AUG-1992.
XX
PF   19-FEB-1991;    91JP-0024633.
XX
PR   20-FEB-1990;    90JP-0040398.
XX
PA   (SHIO ) SHIONOGI & CO LTD.
XX
DR   WPI; 1992-304938/37.
DR   N-PSDB; AAQ27987.
XX
PT   Novel protease prepd. using Bacillus or Saccharomyces host -
PT   capable of cleaving peptide bond at carboxyl terminus of glutamic
PT   acid residues in polypeptide(s)
XX
PS   Disclosure; Page 13-15; 25pp; Japanese.
XX
CC   The sequences given in AAR26842 and AAR29644 are proteases which were
CC   isolated from Staphylococcus aureus strains.  The DNA sequences
CC   encoding these proteins were isolated by PCR using the primer
CC   sequences given in AAQ27960-86.  The protease specifically cleaves
CC   the peptide bond at the C-terminus of the glutamic acid residue in
CC   polypeptide.
XX
SQ   Sequence   357 AA;

Alignment Scores:
Pred. No.:               0.516         Length:        357
Score:                   10.00         Matches:       10
Percent Similarity:      100.00%       Conservative:  0
Best Local Similarity:   100.00%       Mismatches:    0
Query Match:             1.40%         Indels:        0
DB:                      13            Gaps:          0

US-10-008-355-1 (1-2139) x AAR26842 (1-357)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db     233 ThrGlyGlyAsnSerGlySerProValPhe 242

RESULT 13
AAR91034
ID   AAR91034 standard; Protein; 392 AA.
XX
AC   AAR91034;
XX
DT   23-MAY-1996  (first entry)
XX
DE   Beta-galactosidase-V8 protease fusion protein.
XX
KW   V8 protease; Staphylococcus aureus; recombinant protein;
KW   fusion protein; beta-galactosidase; Escherichia coli.
XX
OS   Chimeric Escherichia coli;.
OS   Chimeric Staphylococcus aureus strain V8 (ATCC 27733).
XX
FH   Key             Location/Qualifiers
FT   Region          1..124
FT                   /note= "beta-galactosidase region"
FT   Region          125..392
FT                   /note= "mature V8 protease including the repeat
FT                   region"
XX
PN   EP700995-A2.
XX
PD   13-MAR-1996.
XX
PF   06-SEP-1995;    95EP-0306235.
XX
PR   07-NOV-1994;    94JP-0296028.
PR   07-SEP-1994;    94JP-0238595.
XX
PA   (SUNR ) SUNTORY LTD.
XX
PI   Ohsuye K,  Yabuta M;
XX
DR   WPI; 1996-141021/15.
XX
PT   Prodn. of recombinant polypeptide(s) - using host cells transformed
PT   with a gene coding for the desired polypeptide fused to a protective
PT   polypeptide
XX
PS   Example 2; Page 14-15; 44pp; English.
XX
CC   Fusion proteins (AAR91033 and AAR91034) were constructed comprising
CC   an Escherichia coli beta-galactosidase derivative (protective
CC   polypeptide) fused at its C-terminal end to the S. aureus mature
CC   V8 protease without or with the repeat region.  The constructs
CC   were inserted into vector pG97S4DhCT(G)R6, yielding pV8RPT(-) and
CC   pV8RPT(+), respectively.  Both constructs yielded active protease
CC   when expressed in E. coli JM101 transformants.
XX
SQ   Sequence   392 AA;

Alignment Scores:
Pred. No.:               0.512         Length:        392
Score:                   10.00         Matches:       10
Percent Similarity:      100.00%       Conservative:  0
Best Local Similarity:   100.00%       Mismatches:    0
Query Match:             1.40%         Indels:        0
DB:                      17            Gaps:          0

US-10-008-355-1 (1-2139) x AAR91034 (1-392)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db     289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 14
AAR91035
ID   AAR91035 standard; Protein; 532 AA.
XX
AC   AAR91035;
XX
DT   23-MAY-1996  (first entry)
XX
DE   Recombinant V8 protease V8D fusion protein.
XX
KW   Linker peptide; V8 protease; Staphylococcus aureus;
KW   recombinant protein; fusion protein; beta-galactosidase;
KW   Escherichia coli; transposon Tn903;
KW   aminoglycoside 3'-phosphotransferase.
XX
OS   Chimeric Escherichia coli;
OS   Chimeric synthetic;
OS   Chimeric Staphylococcus aureus strain V8 (ATCC 27733);
OS   Chimeric transposon Tn903.
XX
FH   Key             Location/Qualifiers
FT   Region          1..100
FT                   /note= "beta-galactosidase region"
FT   Region          101..120
FT                   /note= "R6 linker"
FT   Cleavage-site   104..105
```

```
FT                   /note= "cleavage site for OmpT protease"
FT   Region          125..335
FT                   /note= "V8 protease region"
FT   Region          336..356
FT                   /note= "R6 linker"
FT   Cleavage-site   339..340
FT                   /note= "cleavage site for OmpT protease"
FT   Region          307..532
FT                   /note= "aminoglycoside 3'phosphotransferase
FT                          region"
XX
PN   EP700995-A2.
XX
PD   13-MAR-1996.
XX
PF   06-SEP-1995;   95EP-0306235.
XX
PR   07-NOV-1994;   94JP-0296028.
PR   07-SEP-1994;   94JP-0238595.
XX
PA   (SUNR ) SUNTORY LTD.
XX
PI   Ohsuye K,  Yabuta M;
XX
DR   WPI; 1996-141021/15.
XX
PT   Prodn. of recombinant polypeptide(s) - using host cells transformed
PT   with a gene coding for the desired polypeptide fused to a protective
PT   polypeptide
XX
PS   Example 3; Page 16-18; 44pp; English.
XX
CC   A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
CC   and C are protective polypeptides (derived from Escherichia coli
CC   beta-galactosidase and and Tn903 aminoglycoside 3'-phosphotransferase,
CC   respectively), B is Staphylococcus aureus mature V8 protease
CC   lacking the C-terminal repeat region, and L is a linker peptide
CC   (AAR91032).  The fusion protein is expressed in inactive form in E.
CC   coli.  It is then recovered, solubilised and cleaved at the linker
CC   peptide regions with a protease intrinsic to the host cells, i.e.
CC   OmpT protease, to allow recovery of V8 protease.
XX
SQ   Sequence   532 AA;

Alignment Scores:
Pred. No.:              0.498        Length:        532
Score:                  10.00        Matches:       10
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            1.40%        Indels:        0
DB:                     17           Gaps:          0

US-10-008-355-1 (1-2139) x AAR91035 (1-532)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 15
AAW22219
ID   AAW22219 standard; Protein; 532 AA.
XX
AC   AAW22219;
XX
DT   11-SEP-1997  (first entry)
XX
DE   Protein encoded by pV8D construct.
XX
KW   Mutant; Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;
KW   truncation; wild type; PCR; polymerase chain reaction; amplification;
KW   proteolytic activity; fusion protein; beta-galactosidase; urea.
XX
OS   Chimeric - Escherichia coli.
OS   Chimeric - Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..100
FT                   /note= "E. coli beta-galactosidase portion"
FT   Region          101..124
FT                   /note= "R6 linker sequence"
FT   Region          125..336
FT                   /note= "truncated S. aureus V8 protease portion"
FT   Region          337..360
FT                   /note= "R6 linker sequence"
FT   Region          361..532
FT                   /note= "aminoglucoside 3'-phosphotransferase portion"
XX
PN   EP745669-A2.
XX
PD   04-DEC-1996.
XX
PF   31-MAY-1996;   96EP-0303939.
XX
PR   02-JUN-1995;   95JP-0170086.
XX
PA   (SUNR ) SUNTORY LTD.
XX
PI   Ohsuye K,  Yabuta M;
XX
DR   WPI; 1997-013693/02.
XX
PT   Staphylococcus aureus V8 protease mutants - with increased
PT   resistance to denaturation
XX
PS   Claim 7; Page 16-17; 42pp; English.
XX
CC   The invention relates to new mutant Staphylococcus aureus V8 proteases
CC   which have enzyme activity even under environmental conditions which
CC   promote protein denaturation.  The mutants are based on 3 truncated V8
CC   proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino
CC   acids from the C-terminal of the wild type protease.  The mutants also
CC   contain amino acid substitutions, especially D44E, N71S and/or R147K.
CC   The protein sequence shown here represents a chimaeric protein
CC   comprising a truncated Staphylococcus aureus V8 protease lacking the
CC   prepro and C-terminal 56 amino acids linked, via a synthetic R6 linker,
CC   downstream of the E. coli beta-galactosidase. Also included downstream of
CC   the V8 protease fragment is a second R6 linker and a fragment of the
CC   aminoglucoside 3'-phosphotransferase protein.  The chimaeric sequence
CC   was generated by restriction digestion and ligation from the V8RPT(-)
CC   sequence (see AAW22218) by using a natural EcoRV site which removed a
CC   further 8 amino acid from the C-terminus.  This truncated V8 protease,
CC   designated V8D, retains its level of activity in the presence of a
CC   higher concentration of protein denaturant e.g. 5 M urea.
XX
SQ   Sequence   532 AA;

Alignment Scores:
Pred. No.:              0.498        Length:        532
Score:                  10.00        Matches:       10
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            1.40%        Indels:        0
DB:                     18           Gaps:          0

US-10-008-355-1 (1-2139) x AAW22219 (1-532)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    289 ThrGlyGlyAsnSerGlySerProValPhe 298


Search completed: May 16, 2003, 13:06:34
Job time : 123.5 secs
```

D

GenCore version 5.1.4_p5_4578

OM nucleic - protein search, using frame_plus_n2p model

```
Run on:          May 16, 2003, 13:13:37 ; Search time 83 Seconds
                                          (without alignments)
                                          4971.283 Million cell updates/sec

Title:           US-10-008-355-1
Perfect score:   712
Sequence:        1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:   OLIGO
                 Xgapop 60.0 , Xgapext 60.0
                 Ygapop 60.0 , Ygapext 60.0
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        362588 seqs, 96450795 residues

Word size:       1

Total number of hits satisfying chosen parameters:       683032

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110429_9845/app_query.fasta_1.2311
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=olin2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10008355_@CGN_1_1_25_@runat_16052003_110429_9845
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :       Published_Applications_AA:*
                 1:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                 2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                 3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                 4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
                 5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                 6:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                 7:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
                 8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                 9:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                 10:  /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
                 11:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                 12:  /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
                 13:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
                 14:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	712	100.0	712	9	US-10-008-355-2 A	Sequence 2, Appli
	2	699	98.2	699	9	US-10-008-355-8	Sequence 8, Appli
	3	52	7.3	52	9	US-10-008-355-3	Sequence 3, Appli
	4	21	2.9	26	9	US-10-008-355-18	Sequence 18, Appl
	5	13	1.8	716	9	US-10-008-355-7 A	Sequence 7, Appli
	6	11	1.5	734	9	US-10-008-355-5	Sequence 5, Appli
	7	10	1.4	10	9	US-10-008-355-25	Sequence 25, Appl
	8	10	1.4	52	9	US-10-008-355-4	Sequence 4, Appli
	9	10	1.4	732	9	US-10-008-355-6	Sequence 6, Appli
	10	9	1.3	9	9	US-10-008-355-26	Sequence 26, Appl
	11	9	1.3	720	9	US-10-008-355-9	Sequence 9, Appli
	12	8	1.1	274	9	US-10-166-087-30	Sequence 30, Appl
c	13	8	1.1	312	9	US-09-510-332-17	Sequence 17, Appl
c	14	8	1.1	312	10	US-09-393-634-51	Sequence 51, Appl
	15	8	1.1	659	10	US-09-815-242-10510	Sequence 10510, A
	16	8	1.1	876	10	US-09-801-368-44	Sequence 44, Appl
c	17	8	1.1	960	8	US-08-910-386A-18	Sequence 18, Appl
	18	8	1.1	969	8	US-08-945-749-3	Sequence 3, Appli
	19	8	1.1	1170	12	US-10-135-687-2	Sequence 2, Appli
	20	8	1.1	1209	12	US-10-135-687-4	Sequence 4, Appli
	21	8	1.1	1210	10	US-09-860-352A-2	Sequence 2, Appli
	22	7	1.0	13	10	US-09-938-315-89	Sequence 89, Appl
c	23	7	1.0	21	10	US-09-864-761-36591	Sequence 36591, A
c	24	7	1.0	22	10	US-09-864-761-38461	Sequence 38461, A
	25	7	1.0	31	10	US-09-938-315-57	Sequence 57, Appl
c	26	7	1.0	34	10	US-09-864-761-39870	Sequence 39870, A
c	27	7	1.0	36	10	US-09-864-761-39443	Sequence 39443, A
	28	7	1.0	39	9	US-09-774-639-266	Sequence 266, App
	29	7	1.0	39	9	US-09-969-730-368	Sequence 368, App
c	30	7	1.0	40	12	US-10-042-417-16	Sequence 16, Appl
	31	7	1.0	44	10	US-09-925-300-1089	Sequence 1089, Ap
	32	7	1.0	60	9	US-10-074-095-324	Sequence 324, App
	33	7	1.0	60	10	US-09-764-860-324	Sequence 324, App
c	34	7	1.0	66	9	US-09-764-891-5217	Sequence 5217, Ap
c	35	7	1.0	68	10	US-09-764-877-1813	Sequence 1813, Ap
c	36	7	1.0	73	10	US-09-764-877-1970	Sequence 1970, Ap
c	37	7	1.0	74	10	US-09-925-301-1441	Sequence 1441, Ap
	38	7	1.0	76	9	US-09-738-626-5670	Sequence 5670, Ap
c	39	7	1.0	92	10	US-09-764-877-1694	Sequence 1694, Ap
c	40	7	1.0	93	10	US-09-925-300-961	Sequence 961, App
c	41	7	1.0	95	9	US-10-116-016-29	Sequence 29, Appl
c	42	7	1.0	95	10	US-09-764-848-29	Sequence 29, Appl
c	43	7	1.0	107	9	US-09-848-798-154	Sequence 154, App
	44	7	1.0	107	10	US-09-864-761-44523	Sequence 44523, A
	45	7	1.0	113	10	US-09-745-763-130	Sequence 130, App

ALIGNMENTS

```
RESULT 1
US-10-008-355-2
; Sequence 2, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 712
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-2

Alignment Scores:
Pred. No.:               0           Length:         712
Score:                   712.00      Matches:        712
Percent Similarity:      100.00%     Conservative:   0
Best Local Similarity:   100.00%     Mismatches:     0
Query Match:             100.00%     Indels:         0
```

```
DB:                    9              Gaps:          0

US-10-008-355-1 (1-2139) x US-10-008-355-2 (1-712)

Qy       1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20

Qy      61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAsp 40

Qy     121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60

Qy     181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 SerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80

Qy     241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100

Qy     301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuPro 120

Qy     361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140

Qy     421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160

Qy     481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180

Qy     541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200

Qy     601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro 220

Qy     661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240

Qy     721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln 260

Qy     781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280

Qy     841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300

Qy     901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320

Qy     961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340

Qy    1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360

Qy    1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380

Qy    1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400

Qy    1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420

Qy    1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSer 440

Qy    1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAsp 460

Qy    1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480

Qy    1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500

Qy    1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520

Qy    1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540

Qy    1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560

Qy    1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580

Qy    1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600

Qy    1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620

Qy    1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640

Qy    1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660

Qy    1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680

Qy    2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGly 700

Qy    2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATC 2136
           ||||||||||||||||||||||||||||||||||||
```

```
Db     701 GlnCysProArgLeuIleGlnGluLeuLysLeuIle 712

RESULT 2
US-10-008-355-8
; Sequence 8, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 699
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-8

Alignment Scores:
Pred. No.:              0              Length:         699
Score:                  699.00         Matches:        699
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            98.17%         Indels:         0
DB:                     9              Gaps:           0

US-10-008-355-1 (1-2139) x US-10-008-355-8 (1-699)

Qy       1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20

Qy      61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAsp 40

Qy     121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60

Qy     181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 SerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80

Qy     241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100

Qy     301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuPro 120

Qy     361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140

Qy     421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160

Qy     481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180

Qy     541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200

Qy     601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro 220

Qy     661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240

Qy     721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln 260

Qy     781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280

Qy     841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300

Qy     901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320

Qy     961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340

Qy    1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360

Qy    1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380

Qy    1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400

Qy    1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420

Qy    1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSer 440

Qy    1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAsp 460

Qy    1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480

Qy    1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500

Qy    1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520

Qy    1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540

Qy    1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560
```

```
Qy   1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580

Qy   1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600

Qy   1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620

Qy   1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640

Qy   1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660

Qy   1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680

Qy   2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGG 2097
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrp 699

RESULT 3
US-10-008-355-3
; Sequence 3, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 52
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-3

Alignment Scores:
Pred. No.:              5.6e-42      Length:        52
Score:                  52.00        Matches:       52
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            7.30%        Indels:        0
DB:                     9            Gaps:          0

US-10-008-355-1 (1-2139) x US-10-008-355-3 (1-52)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCT 1989
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAla 20

Qy   1990 TTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGC 2049
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArg 40

Qy   2050 ACAATCAGCGTGGACATCCGCTACGTTCTCTTCATG 2085
          ||||||||||||||||||||||||||||||||||||
Db     41 ThrIleSerValAspIleArgTyrValLeuPheMet 52

RESULT 4
US-10-008-355-18
; Sequence 18, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
;   LENGTH: 26
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-18

Alignment Scores:
Pred. No.:              1.86e-11     Length:        26
Score:                  21.00        Matches:       21
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            2.95%        Indels:        0
DB:                     9            Gaps:          0

US-10-008-355-1 (1-2139) x US-10-008-355-18 (1-26)

Qy     82 ATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTT 141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 MetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPhe 25

Qy    142 ACG 144
          |||
Db     26 Thr 26

RESULT 5
US-10-008-355-7
; Sequence 7, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Xylella fastidiosa
US-10-008-355-7

Alignment Scores:
Pred. No.:              0.000938     Length:        716
Score:                  13.00        Matches:       13
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            1.83%        Indels:        0
DB:                     9            Gaps:          0

US-10-008-355-1 (1-2139) x US-10-008-355-7 (1-716)

Qy    643 GACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGC 681
          |||||||||||||||||||||||||||||||||||||||
```

```
Db     212 AspAsnTrpMetTrpProArgHisThrGlyAspPheSer 224

RESULT 6
US-10-008-355-5
; Sequence 5, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
;   LENGTH: 734
;   TYPE: PRT
;   ORGANISM: Shewanella putrefaciens
US-10-008-355-5

Alignment Scores:
Pred. No.:               0.0867        Length:         734
Score:                   11.00         Matches:        11
Percent Similarity:      100.00%       Conservative:   0
Best Local Similarity:   100.00%       Mismatches:     0
Query Match:             1.54%         Indels:         0
DB:                      9             Gaps:           0

US-10-008-355-1 (1-2139) x US-10-008-355-5 (1-734)

Qy     643 GACAACTGGATGTGGCCGCGTCACACGGGCGAC 675
           |||||||||||||||||||||||||||||||||
Db     222 AspAsnTrpMetTrpProArgHisThrGlyAsp 232

RESULT 7
US-10-008-355-25
; Sequence 25, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
;   LENGTH: 10
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Consensus sequence
US-10-008-355-25

Alignment Scores:
Pred. No.:               1.37          Length:         10
Score:                   10.00         Matches:        10
Percent Similarity:      100.00%       Conservative:   0
Best Local Similarity:   100.00%       Mismatches:     0
Query Match:             1.40%         Indels:         0
DB:                      9             Gaps:           0

US-10-008-355-1 (1-2139) x US-10-008-355-25 (1-10)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db       1 ThrGlyGlyAsnSerGlySerProValPhe 10

RESULT 8
US-10-008-355-4
; Sequence 4, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 52
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-10-008-355-4

Alignment Scores:
Pred. No.:               1.13          Length:         52
Score:                   10.00         Matches:        10
Percent Similarity:      100.00%       Conservative:   0
Best Local Similarity:   100.00%       Mismatches:     0
Query Match:             1.40%         Indels:         0
DB:                      9             Gaps:           0

US-10-008-355-1 (1-2139) x US-10-008-355-4 (1-52)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
Db       1 ThrGlyGlyAsnSerGlySerProValPhe 10

RESULT 9
US-10-008-355-6
; Sequence 6, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
;   LENGTH: 732
;   TYPE: PRT
;   ORGANISM: Shewanella putrefaciens
US-10-008-355-6

Alignment Scores:
Pred. No.:               0.835         Length:         732
Score:                   10.00         Matches:        10
Percent Similarity:      100.00%       Conservative:   0
Best Local Similarity:   100.00%       Mismatches:     0
Query Match:             1.40%         Indels:         0
DB:                      9             Gaps:           0

US-10-008-355-1 (1-2139) x US-10-008-355-6 (1-732)

Qy    1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
           ||||||||||||||||||||||||||||||
```

```
Db    662 ThrGlyGlyAsnSerGlySerProValPhe 671

RESULT 10
US-10-008-355-26
; Sequence 26, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Consensus sequence
US-10-008-355-26

Alignment Scores:
Pred. No.:              2.14e+07     Length:         9
Score:                  9.00         Matches:        9
Percent Similarity:     100.00%      Conservative:   0
Best Local Similarity:  100.00%      Mismatches:     0
Query Match:            1.26%        Indels:         0
DB:                     9            Gaps:           0

US-10-008-355-1 (1-2139) x US-10-008-355-26 (1-9)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTA 1956
          |||||||||||||||||||||||||||
Db      1 ThrGlyGlyAsnSerGlySerProVal 9

RESULT 11
US-10-008-355-9
; Sequence 9, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-9

Alignment Scores:
Pred. No.:              8.05         Length:         720
Score:                  9.00         Matches:        9
Percent Similarity:     100.00%      Conservative:   0
Best Local Similarity:  100.00%      Mismatches:     0
Query Match:            1.26%        Indels:         0
DB:                     9            Gaps:           0

US-10-008-355-1 (1-2139) x US-10-008-355-9 (1-720)

Qy    655 TGGCCGCGTCACACGGGCGACTTCAGC 681
          |||||||||||||||||||||||||||
Db    221 TrpProArgHisThrGlyAspPheSer 229

RESULT 12
US-10-166-087-30
; Sequence 30, Application US/10166087
; Publication No. US20030077767A1
; GENERAL INFORMATION:
;  APPLICANT: Ecopia Biosciences Inc.
;  APPLICANT:  Farnet, Chris
;  APPLICANT:  Staffa, Alfredo
;  APPLICANT:  Zazopoulos, Emmanuel
;  TITLE OF INVENTION: Genes and proteins for the biosynthesis of anthramycin
;  FILE REFERENCE: 3014-2US.
;  CURRENT APPLICATION NUMBER: US/10/166,087
;  CURRENT FILING DATE:  2002-06-11
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
;   LENGTH: 274
;   TYPE: PRT
;   ORGANISM: Streptomyces refuineus subspecies thermotolerans
US-10-166-087-30

Alignment Scores:
Pred. No.:              86.6         Length:         274
Score:                  8.00         Matches:        8
Percent Similarity:     100.00%      Conservative:   0
Best Local Similarity:  100.00%      Mismatches:     0
Query Match:            1.12%        Indels:         0
DB:                     9            Gaps:           0

US-10-008-355-1 (1-2139) x US-10-166-087-30 (1-274)

Qy    121 CGAATGCGTGAGCTCGGCTTTACG 144
          ||||||||||||||||||||||||
Db    168 ArgMetArgGluLeuGlyPheThr 175

RESULT 13
US-09-510-332-17
; Sequence 17, Application US/09510332
; Publication No. US20030022278A1
; GENERAL INFORMATION:
;  APPLICANT:  Zuker, Charles S.
;  APPLICANT:  Adler, Jon Elliot
;  APPLICANT:  Ryba, Nick
;  APPLICANT:  Mueller, Ken
;  APPLICANT:  Hoon, Mark
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION:  T2R, a No. US20030022278A1el Family of Taste Receptors
;  FILE REFERENCE:  02307E-098010US
;  CURRENT APPLICATION NUMBER: US/09/510,332
;  CURRENT FILING DATE:  2000-02-22
;  PRIOR APPLICATION NUMBER:  US 09/393,634
;  PRIOR FILING DATE:  1999-09-10
;  NUMBER OF SEQ ID NOS:  172
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 17
;   LENGTH: 312
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: human T2R09 (hGR09)
US-09-510-332-17

Alignment Scores:
Pred. No.:              85.3         Length:         312
Score:                  8.00         Matches:        8
Percent Similarity:     100.00%      Conservative:   0
Best Local Similarity:  100.00%      Mismatches:     0
Query Match:            1.13%        Indels:         0
DB:                     9            Gaps:           0
```

```
US-10-008-355-1 (1-2139) x US-09-510-332-17 (1-312)

Qy    1164 CTTGGCTCCTTCCTTATAAGCCTT 1141
           ||||||||||||||||||||||||
Db     135 LeuGlySerPheLeuIleSerLeu 142

RESULT 14
US-09-393-634-51
; Sequence 51, Application US/09393634
; Patent No. US20020051997A1
; GENERAL INFORMATION:
;  APPLICANT:  Zuker, Charles S.
;  APPLICANT:  Adler, Jon Elliot
;  APPLICANT:  Ryba, Nick
;  APPLICANT:  Mueller, Ken
;  APPLICANT:  Hoon, Mark
;  APPLICANT:  The Regents of the University of California
;  APPLICANT:  The Government of the United States of America
;  APPLICANT:  as represented by the Secretary of the
;  APPLICANT:  Department of Health and Human Services
;  TITLE OF INVENTION:  SF, a No. US20020051997A1el Family of Taste Receptors
;  FILE REFERENCE:  02307E-098000US
;  CURRENT APPLICATION NUMBER: US/09/393,634
;  CURRENT FILING DATE:  1999-09-10
;  NUMBER OF SEQ ID NOS:  92
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 51
;   LENGTH: 312
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: human GR09
US-09-393-634-51

Alignment Scores:
Pred. No.:              85.3        Length:        312
Score:                  8.00        Matches:       8
Percent Similarity:     100.00%     Conservative:  0
Best Local Similarity:  100.00%     Mismatches:    0
Query Match:            1.13%       Indels:        0
DB:                     10          Gaps:          0

US-10-008-355-1 (1-2139) x US-09-393-634-51 (1-312)

Qy    1164 CTTGGCTCCTTCCTTATAAGCCTT 1141
           ||||||||||||||||||||||||
Db     135 LeuGlySerPheLeuIleSerLeu 142

RESULT 15
US-09-815-242-10510
; Sequence 10510, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10510
;   LENGTH: 659
;   TYPE: PRT
;   ORGANISM: Enterococcus faecalis
US-09-815-242-10510

Alignment Scores:
Pred. No.:              78.3        Length:        659
Score:                  8.00        Matches:       8
Percent Similarity:     100.00%     Conservative:  0
Best Local Similarity:  100.00%     Mismatches:    0
Query Match:            1.12%       Indels:        0
DB:                     10          Gaps:          0

US-10-008-355-1 (1-2139) x US-09-815-242-10510 (1-659)

Qy     490 GCCAAAAAAGAAAATGCAGACGAG 513
           ||||||||||||||||||||||||
Db     206 AlaLysLysGluAsnAlaAspGlu 213


Search completed: May 16, 2003, 13:41:27
Job time : 98 secs
```

D
P

GenCore version 5.1.4_p5_4578

OM nucleic - protein search, using frame_plus_n2p model

```
Run on:          May 16, 2003, 13:00:32 ; Search time 79 Seconds
                                          (without alignments)
                                          5205.861 Million cell updates/sec

Title:           US-10-008-355-1
Perfect score:   712
Sequence:        1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:   OLIGO
                 Xgapop 60.0 , Xgapext 60.0
                 Ygapop 60.0 , Ygapext 60.0
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        283224 seqs, 96134422 residues

Word size:       1

Total number of hits satisfying chosen parameters:      565536

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110426_9746/app_query.fasta_1.2311
-DB=PIR_73 -QFMT=fastan -SUFFIX=olin2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_111_@runat_16052003_110426_9746 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :       PIR_73:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	13	1.8	716	2	G82627	hypothetical prote
	2	10	1.4	336	1	PRSASK	glutamyl endopepti
	3	10	1.4	342	2	G89873	hypothetical prote
	4	10	1.4	357	2	S21758	glutamic acid-spec
	5	10	1.4	569	2	F75381	probable two-compo
	6	9	1.3	266	2	D75483	hypothetical prote
	7	9	1.3	452	2	AE1710	hypothetical prote
	8	9	1.3	483	2	AF2363	hypothetical prote
	9	9	1.3	684	2	T36771	probable integral
	10	9	1.3	984	1	A34076	protein-tyrosine k
	11	9	1.3	1116	2	B70476	hypothetical prote
	12	8	1.1	102	2	T36095	hypothetical prote
	13	8	1.1	106	2	T04804	hypothetical prote
	14	8	1.1	116	2	JC4669	cocaine-and amphet
c	15	8	1.1	162	2	C69495	conserved hypothet
	16	8	1.1	189	2	A69054	hypothetical prote
	17	8	1.1	214	2	C84259	triosephosphate is
c	18	8	1.1	224	2	A81449	probable molybdenu
c	19	8	1.1	226	2	S04752	H+-transporting tw
c	20	8	1.1	241	2	A69299	ABC transporter, A
	21	8	1.1	256	2	A34770	ORF1 protein - sai
	22	8	1.1	284	2	T07614	homeobox-leucine z
	23	8	1.1	288	2	F72219	conserved hypothet
c	24	8	1.1	332	2	S46947	ribosomal protein
	25	8	1.1	349	2	S40718	hypothetical prote
	26	8	1.1	382	2	T39451	methionine adenosy
	27	8	1.1	392	2	E83899	thiolase (acetyl-C
	28	8	1.1	400	2	E96029	probable beta-keto
	29	8	1.1	401	2	AI3115	beta-ketoadipyl Co
	30	8	1.1	419	2	D98171	beta-ketoadipyl-Co
c	31	8	1.1	421	2	C91109	integrase [importe
c	32	8	1.1	421	2	F85954	probable pathogeni
	33	8	1.1	425	2	S54010	1-aminocyclopropan
	34	8	1.1	444	2	S54011	1-aminocyclopropan
	35	8	1.1	458	2	F75367	potassium uptake p
c	36	8	1.1	479	2	T44326	hypothetical prote
c	37	8	1.1	513	2	T34689	probable iron-sulf
	38	8	1.1	544	2	D64882	periplasmic oligop
	39	8	1.1	544	2	G90867	probable transport
	40	8	1.1	544	2	B85751	probable transport
	41	8	1.1	563	2	B97319	homolog of plant a
c	42	8	1.1	570	2	A48836	fibropellin C prec
c	43	8	1.1	586	2	S71235	chaperonin 60 alph
c	44	8	1.1	633	2	G95385	Kup2 Potassium upt
	45	8	1.1	637	1	YCRP	acetolactate synth

ALIGNMENTS

```
RESULT 1
G82627
hypothetical protein XF1887 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: G82627
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: G82627
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-716 <SIM>
A;Cross-references: GB:AE004008; GB:AE003849; NID:g9106961; PIDN:AAF84693.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1887

Alignment Scores:
Pred. No.:            0.000396     Length:         716
Score:                13.00        Matches:        13
Percent Similarity:   100.00%      Conservative:   0
```

39 102a
'00

```
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.83%          Indels:        0
DB:                     2              Gaps:          0

US-10-008-355-1 (1-2139) x G82627 (1-716)

Qy    643 GACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGC 681
          |||||||||||||||||||||||||||||||||||||||
Db    212 AspAsnTrpMetTrpProArgHisThrGlyAspPheSer 224
```

681
-643
38

RESULT 2
PRSASK
glutamyl endopeptidase (EC 3.4.21.19) precursor - Staphylococcus aureus
N;Alternate names: staphylococcal serine proteinase
C;Species: Staphylococcus aureus
C;Date: 04-Dec-1986 #sequence_revision 30-Jun-1991 #text_change 18-Jun-1999
C;Accession: A26812; A00966
R;Carmona, C.; Gray, G.L.
Nucleic Acids Res. 15, 6757, 1987
A;Title: Nucleotide sequence of the serine protease gene of Staphylococcus aureus, strai
A;Reference number: A26812; MUID:87316953; PMID:3306605
A;Accession: A26812
A;Molecule type: DNA
A;Residues: 1-336 <CAR>
A;Cross-references: EMBL:Y00356; NID:g46686; PIDN:CAA68434.1; PID:g46687
A;Experimental source: strain V8
R;Drapeau, G.R.
Can. J. Biochem. 56, 534-544, 1978
A;Title: The primary structure of staphylococcal protease.
A;Reference number: A23824; MUID:78212487; PMID:96922
A;Accession: A00966
A;Molecule type: protein
A;Residues: 69-108;110-124;126-144,'D',146-192,'T',194-228,'N',230-258,'Q',260,'D',262-2
A;Experimental source: strain V8
C;Comment: This extracellular proteolytic enzyme preferentially cleaves the peptide bond
C;Comment: This enzyme may be distantly related to the trypsin-type serine proteinases,
dues.
C;Superfamily: staphylococcal serine proteinase
C;Keywords: hydrolase; serine proteinase
F;69-336/Product: staphylococcal serine proteinase #status experimental <ACT>
F;119,161/Active site: His, Asp #status predicted
F;237/Active site: Ser #status experimental

```
Alignment Scores:
Pred. No.:              0.513          Length:        336
Score:                  10.00          Matches:       10
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.40%          Indels:        0
DB:                     1              Gaps:          0
```

.91 38

```
US-10-008-355-1 (1-2139) x PRSASK (1-336)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    233 ThrGlyGlyAsnSerGlySerProValPhe 242
```

RESULT 3
G89873
hypothetical protein sspA [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: G89873
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: G89873
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-342 <KUR>
A;Cross-references: GB:BA000018; PID:g13700850; PIDN:BAB42146.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: sspA
C;Superfamily: staphylococcal serine proteinase

```
Alignment Scores:
Pred. No.:              0.512          Length:        342
Score:                  10.00          Matches:       10
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.40%          Indels:        0
DB:                     2              Gaps:          0

US-10-008-355-1 (1-2139) x G89873 (1-342)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    233 ThrGlyGlyAsnSerGlySerProValPhe 242
```

RESULT 4
S21758
glutamic acid-specific endopeptidase - Staphylococcus aureus
C;Species: Staphylococcus aureus
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S21758
R;Yoshikawa, K.; Tsuzuki, H.; Fujiwara, T.; Nakamura, E.; Iwamoto, H.; Matsumoto, K.;
Biochim. Biophys. Acta 1121, 221-228, 1992
A;Title: Purification, characterization and gene cloning of a novel glutamic acid-spe
A;Reference number: S21758; MUID:92287954; PMID:1599945
A;Accession: S21758
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-357 <YOS>
A;Cross-references: GB:D00730; NID:g216970; PIDN:BAA00630.1; PID:g216971
C;Superfamily: staphylococcal serine proteinase

```
Alignment Scores:
Pred. No.:              0.51           Length:        357
Score:                  10.00          Matches:       10
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.40%          Indels:        0
DB:                     2              Gaps:          0

US-10-008-355-1 (1-2139) x S21758 (1-357)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    233 ThrGlyGlyAsnSerGlySerProValPhe 242
```

RESULT 5
F75381
probable two-component sensor - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: F75381
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75381
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-569 <WHI>
A;Cross-references: GB:AE001999; GB:AE000513; NID:g6459316; PIDN:AAF11119.1; PID:g645
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1556
A;Map position: 1

```
Alignment Scores:
Pred. No.:              0.484       Length:        569
Score:                  10.00       Matches:       10
Percent Similarity:     100.00%     Conservative:  0
Best Local Similarity:  100.00%     Mismatches:    0
Query Match:            1.40%       Indels:        0
DB:                     2           Gaps:          0

US-10-008-355-1 (1-2139) x F75381 (1-569)

Qy      25 CTTCTCGGAGCAGCCCTGCTGTTGGGTGCT 54
           ||||||||||||||||||||||||||||||
Db      63 LeuLeuGlyAlaAlaLeuLeuLeuGlyAla 72

RESULT 6
D75483
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
C;Accession: D75483
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: D75483
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-266 <WHI>
A;Cross-references: GB:AE001928; GB:AE000513; NID:g6458421; PIDN:AAF10302.1; PID:g645843
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0714
A;Map position: 1
C;Superfamily: Deinococcus radiodurans hypothetical protein DR0714

Alignment Scores:
Pred. No.:              5.59        Length:        266
Score:                  9.00        Matches:       9
Percent Similarity:     100.00%     Conservative:  0
Best Local Similarity:  100.00%     Mismatches:    0
Query Match:            1.26%       Indels:        0
DB:                     2           Gaps:          0

US-10-008-355-1 (1-2139) x D75483 (1-266)

Qy      25 CTTCTCGGAGCAGCCCTGCTGTTGGGT 51
           |||||||||||||||||||||||||||
Db     108 LeuLeuGlyAlaAlaLeuLeuLeuGly 116

RESULT 7
AE1710
hypothetical protein homolog lin2224 [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AE1710
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1710
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-452 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC97453.1; PID:g16414737; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin2224

Alignment Scores:
Pred. No.:              5.27        Length:        452
Score:                  9.00        Matches:       9
Percent Similarity:     100.00%     Conservative:  0
Best Local Similarity:  100.00%     Mismatches:    0
Query Match:            1.26%       Indels:        0
DB:                     2           Gaps:          0

US-10-008-355-1 (1-2139) x AE1710 (1-452)

Qy    1555 GTAGAGCTTTCCAAGAGCGTAATAGCT 1581
           |||||||||||||||||||||||||||
Db     142 ValGluLeuSerLysSerValIleAla 150

RESULT 8
AF2363
hypothetical protein all4462 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AF2363
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AF2363
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-483 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB76161.1; PID:g17133598; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all4462

Alignment Scores:
Pred. No.:              5.23        Length:        483
Score:                  9.00        Matches:       9
Percent Similarity:     100.00%     Conservative:  0
Best Local Similarity:  100.00%     Mismatches:    0
Query Match:            1.26%       Indels:        0
DB:                     2           Gaps:          0

US-10-008-355-1 (1-2139) x AF2363 (1-483)

Qy      34 GCAGCCCTGCTGTTGGGTGCTTCAGGG 60
           |||||||||||||||||||||||||||
Db     353 AlaAlaLeuLeuLeuGlyAlaSerGly 361

RESULT 9
T36771
probable integral membrane protein - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36771
R;Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A
submitted to the EMBL Data Library, July 1999
A;Reference number: Z21574
A;Accession: T36771
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-684 <SEE>
A;Cross-references: EMBL:AL096844; PIDN:CAB50875.1; GSPDB:GN00070; SCOEDB:SCI28.01
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCI28.01

Alignment Scores:
Pred. No.:              5.03        Length:        684
Score:                  9.00        Matches:       9
Percent Similarity:     100.00%     Conservative:  0
Best Local Similarity:  100.00%     Mismatches:    0
```

```
Query Match:            1.26%          Indels:        0
DB:                     2              Gaps:          0

US-10-008-355-1 (1-2139) x T36771 (1-684)

Qy      28 CTCGGAGCAGCCCTGCTGTTGGGTGCT 54
           |||||||||||||||||||||||||||
Db     442 LeuGlyAlaAlaLeuLeuLeuGlyAla 450

RESULT 10
A34076
protein-tyrosine kinase (EC 2.7.1.112) receptor type eph 1 precursor - human
N;Alternate names: receptor tyrosine kinase eph
C;Species: Homo sapiens (man)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 04-Feb-2000
C;Accession: A34076; S44280
R;Hirai, H.; Maru, Y.; Hagiwara, K.; Nishida, J.; Takaku, F.
Science 238, 1717-1720, 1987
A;Title: A novel putative tyrosine kinase receptor encoded by the eph gene.
A;Reference number: A34076; MUID:88070650; PMID:2825356
A;Accession: A34076
A;Molecule type: mRNA
A;Residues: 1-984 <HIR>
A;Cross-references: GB:M18391; NID:g339716; PIDN:AAA36747.1; PID:g339717
A;Note: the sequence in GenBank entry HUMTKR, release 111.0, has the codons GCG for 398-
R;Tuzi, N.L.
submitted to the EMBL Data Library, November 1993
A;Description: An EGFR/eph chimeric receptor possesses ligand stimulated tyrosine kinase
A;Reference number: S44280
A;Accession: S44280
A;Molecule type: mRNA
A;Residues: 286-397,'A',399-580,'QRDRATDVDREDKLWLKPYVDLQAYEDPAQGALDF',583,625-984 <TUZ>
A;Cross-references: EMBL:Z27409; NID:g482916; PIDN:CAA81796.1; PID:g482917
C;Genetics:
A;Gene: GDB:EPHT1; EPH; EPHT
A;Cross-references: GDB:119875; OMIM:179610
A;Map position: 7q32-7q36
C;Superfamily: protein-tyrosine kinase, receptor type eph; fibronectin type III repeat h
C;Keywords: ATP; autophosphorylation; glycoprotein; kinase-related transforming protein;
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-984/Product: protein-tyrosine kinase receptor type eph 1 #status predicted <MAT>
F;548-568/Domain: transmembrane #status predicted <TMM>
F;630-895/Domain: protein kinase homology <KIN>
F;638-646/Region: protein kinase ATP-binding motif
F;918-984/Domain: SAM homology <SAM>
F;59,338,414,478/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.:              4.83           Length:        984
Score:                  9.00           Matches:       9
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.26%          Indels:        0
DB:                     1              Gaps:          0

US-10-008-355-1 (1-2139) x A34076 (1-984)

Qy      25 CTTCTCGGAGCAGCCCTGCTGTTGGGT 51
           |||||||||||||||||||||||||||
Db     556 LeuLeuGlyAlaAlaLeuLeuLeuGly 564

RESULT 11
B70476
hypothetical protein aq_2054 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C;Accession: B70476
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70476
```

```
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1116 <AQF>
A;Cross-references: GB:AE000770; NID:g2984274; PIDN:AAC07805.1; PID:g2984280; GB:AE00
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_2054

Alignment Scores:
Pred. No.:              4.76           Length:        1116
Score:                  9.00           Matches:       9
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.26%          Indels:        0
DB:                     2              Gaps:          0

US-10-008-355-1 (1-2139) x B70476 (1-1116)

Qy    1405 GTAATCGACAAGAAATTCAAAGGCGAC 1431
           |||||||||||||||||||||||||||
Db     329 ValIleAspLysLysPheLysGlyAsp 337

RESULT 12
T36095
hypothetical protein SCE134.13 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
C;Accession: T36095
R;Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M
submitted to the EMBL Data Library, April 1999
A;Reference number: Z21596
A;Accession: T36095
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-102 <SAU>
A;Cross-references: EMBL:AL049661; PIDN:CAB41211.1; GSPDB:GN00070; SCOEDB:SCE134.13
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE134.13
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE134.13

Alignment Scores:
Pred. No.:              65.9           Length:        102
Score:                  8.00           Matches:       8
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.12%          Indels:        0
DB:                     2              Gaps:          0

US-10-008-355-1 (1-2139) x T36095 (1-102)

Qy      31 GGAGCAGCCCTGCTGTTGGGTGCT 54
           ||||||||||||||||||||||||
Db      79 GlyAlaAlaLeuLeuLeuGlyAla 86

RESULT 13
T04804
hypothetical protein F10M23.150 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T04804
R;Bevan, M.; Lecharny, A.; Chefdor, F.; Krivitzky, M.; Kreis, M.; Hoheisel, J.; Mewes
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15385
A;Accession: T04804
A;Molecule type: DNA
A;Residues: 1-106 <BEV>
A;Cross-references: EMBL:AL035440
A;Experimental source: cultivar Columbia; BAC clone F10M23
C;Genetics:
A;Map position: 4
A;Introns: 58/3
A;Note: F10M23.150
```

```
Alignment Scores:
Pred. No.:              65.7          Length:        106
Score:                  8.00          Matches:       8
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.12%         Indels:        0
DB:                     2             Gaps:          0

US-10-008-355-1 (1-2139) x T04804 (1-106)

Qy      10 AAATTAAAAAGTATTCTTCTCGGA 33
           ||||||||||||||||||||||||
Db      71 LysLeuLysSerIleLeuLeuGly 78

RESULT 14
JC4669
cocaine-and amphetamine regulated transcript protein - human
C;Species: Homo sapiens (man)
C;Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 21-Jul-2000
C;Accession: JC4669; G01648
R;Douglass, J.; Daoud, S.
Gene 169, 241-245, 1996
A;Title: Characterization of the human cDNA and genomic DNA encoding CART: A cocaine- an
A;Reference number: JC4669; MUID:96194810; PMID:8647455
A;Accession: JC4669
A;Molecule type: mRNA
A;Residues: 1-116 <DOU>
A;Cross-references: GB:U16826; NID:g609305; PIDN:AAB08010.1; PID:g609306
R;Douglass, J.O.
submitted to the EMBL Data Library, January 1995
A;Reference number: G08029
A;Accession: G01648
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-116 <DO2>
A;Cross-references: EMBL:U20325; NID:g665578; PID:g665579
C;Comment: This protein plays a conserved functional role within the mammalian central r
stimulants.
C;Genetics:
A;Gene: CART
A;Map position: 5
A;Introns: 53/3; 81/3

Alignment Scores:
Pred. No.:              65            Length:        116
Score:                  8.00          Matches:       8
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.12%         Indels:        0
DB:                     2             Gaps:          0

US-10-008-355-1 (1-2139) x JC4669 (1-116)

Qy      25 CTTCTCGGAGCAGCCCTGCTGTTG 48
           ||||||||||||||||||||||||
Db      11 LeuLeuGlyAlaAlaLeuLeuLeu 18

RESULT 15
C69495
conserved hypothetical protein AF1964 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: C69495
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: C69495
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-162 <KLE>
A;Cross-references: GB:AE000967; GB:AE000782; NID:g2689290; PIDN:AAB89289.1; PID:g264

Alignment Scores:
Pred. No.:              62.6          Length:        162
Score:                  8.00          Matches:       8
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.13%         Indels:        0
DB:                     2             Gaps:          0

US-10-008-355-1 (1-2139) x C69495 (1-162)

Qy    1401 CTTGAATATATCGGGGAGCTTGTC 1378
           ||||||||||||||||||||||||
Db     142 LeuGluTyrIleGlyGluLeuVal 149


Search completed: May 16, 2003, 13:16:20
Job time : 92 secs
```

```
                         GenCore version 5.1.4_p5_4578
               Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:         May 16, 2003, 12:55:27 ; Search time 165 Seconds
                                          (without alignments)
                                          5342.240 Million cell updates/sec

Title:          US-10-008-355-1
Perfect score:  712
Sequence:       1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:  OLIGO
                Xgapop 60.0 , Xgapext 60.0
                Ygapop 60.0 , Ygapext 60.0
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       671580 seqs, 206047115 residues

Word size:      1

Total number of hits satisfying chosen parameters:      1343044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110426_9734/app_query.fasta_1.2311
-DB=SPTREMBL_21 -QFMT=fastan -SUFFIX=olin2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_238_@runat_16052003_110426_9734 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :      SPTREMBL_21:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                    SUMMARIES
                        %
Result               Query
  No.      Score  Match Length DB  ID                    Description
-------------------------------------------------------------------------------
    1        13    1.8    716  16  Q9PC94                Q9pc94 xylella fas
    2        10    1.4    316   2  Q9FBG1                Q9fbg1 staphylococ
    3        10    1.4    342  16  Q99V45                Q99v45 staphylococ
    4        10    1.4    357   2  Q04186                Q04186 staphylococ
    5        10    1.4    569  16  Q9RU39                Q9ru39 deinococcus
    6         9    1.3    217   2  Q9AJX0                Q9ajx0 staphylococ
    7         9    1.3    266  16  Q9RWF5                Q9rwf5 deinococcus
    8         9    1.3    452  16  Q929Q0                Q929q0 listeria in
    9         9    1.3    483  16  Q8YNU9                Q8ynu9 anabaena sp
   10         9    1.3    684  16  Q9S282                Q9s282 streptomyce
   11         8    1.1     92  11  Q91WB5                Q91wb5 mus musculu
   12         8    1.1    102  16  Q9X868                Q9x868 streptomyce
   13         8    1.1    106  10  Q9SZ21                Q9sz21 arabidopsis
c  14         8    1.1    162  17  O28315                O28315 archaeoglob
   15         8    1.1    177  17  Q8ZVA1                Q8zva1 pyrobaculum
   16         8    1.1    178  16  Q9K3J5                Q9k3j5 streptomyce
   17         8    1.1    189  17  O27455                O27455 methanobact
   18         8    1.1    199  10  Q9SS66                Q9ss66 arabidopsis
c  19         8    1.1    211   5  Q9V4H5                Q9v4h5 drosophila
c  20         8    1.1    224  16  Q9PIJ8                Q9pij8 campylobact
   21         8    1.1    226   4  Q8WZ66                Q8wz66 homo sapien
c  22         8    1.1    241  17  O29854                O29854 archaeoglob
c  23         8    1.1    268  17  Q8TV91                Q8tv91 methanopyru
   24         8    1.1    284  10  Q39862                Q39862 glycine max
   25         8    1.1    288  16  Q9X246                Q9x246 thermotoga
c  26         8    1.1    312   4  Q9NYW1                Q9nyw1 homo sapien
c  27         8    1.1    332  10  Q40600                Q40600 oenothera b
c  28         8    1.1    344   2  Q93CA1                Q93ca1 bifidobacte
   29         8    1.1    350   9  Q9ZXM6                Q9zxm6 bacteriopha
   30         8    1.1    356   2  O33999                O33999 chromatium
   31         8    1.1    372   5  Q9NGT6                Q9ngt6 leishmania
   32         8    1.1    380   2  Q9FDK6                Q9fdk6 zymomonas m
   33         8    1.1    391   5  Q93230                Q93230 caenorhabdi
   34         8    1.1    392  16  Q9KBD1                Q9kbd1 bacillus ha
c  35         8    1.1    395   2  Q9AL66                Q9al66 shigella fl
c  36         8    1.1    396   2  Q8VNQ8                Q8vnq8 escherichia
   37         8    1.1    396   4  Q96RL8                Q96rl8 homo sapien
   38         8    1.1    396   4  Q8WWV3                Q8wwv3 homo sapien
   39         8    1.1    400   2  Q8VV03                Q8vv03 streptomyce
   40         8    1.1    400  16  Q92TL0                Q92tl0 rhizobium m
c  41         8    1.1    414  16  Q9F2U9                Q9f2u9 streptomyce
   42         8    1.1    419  16  Q8U7A5                Q8u7a5 agrobacteri
c  43         8    1.1    421   2  Q9ADX1                Q9adx1 escherichia
c  44         8    1.1    421   2  Q8VQN9                Q8vqn9 escherichia
c  45         8    1.1    421   2  Q8VLM4                Q8vlm4 escherichia

                              ALIGNMENTS

RESULT 1
Q9PC94
ID   Q9PC94      PRELIMINARY;      PRT;   716 AA.
AC   Q9PC94;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Xf1887.
GN   XF1887.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
```

```
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004008; AAF84693.1; -.
DR   MEROPS; S46.001; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   716 AA;  79375 MW;  EAF086E2315BBDFC CRC64;

Alignment Scores:
Pred. No.:              0.000331      Length:        716
Score:                  13.00         Matches:       13
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.83%         Indels:        0
DB:                     16            Gaps:          0

US-10-008-355-1 (1-2139) x Q9PC94 (1-716)

Qy    643 GACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGC 681
          |||||||||||||||||||||||||||||||||||||||
Db    212 AspAsnTrpMetTrpProArgHisThrGlyAspPheSer 224

RESULT 2
Q9FBG1
ID   Q9FBG1     PRELIMINARY;      PRT;   316 AA.
AC   Q9FBG1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Glutamyl endopeptidase.
GN   PROM.
OS   Staphylococcus warneri.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=1292;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kakikawa M.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kodaira K.I.;
RT   "Characterization of the gene encoding glutamyl endopeptidase of
RT   Staphylococcus warneri M.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ293885; CAC06168.1; -.
DR   MEROPS; S01.269; -.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   InterPro; IPR000126; Ser_proteas_V8.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00839; V8PROTEASE.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00673; V8_SER; 1.
KW   Hydrolase; Serine protease.
SQ   SEQUENCE   316 AA;  34296 MW;  4E997A5A111DDB40 CRC64;

Alignment Scores:
Pred. No.:              0.599         Length:        316
Score:                  10.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.40%         Indels:        0
DB:                     2             Gaps:          0

US-10-008-355-1 (1-2139) x Q9FBG1 (1-316)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    231 ThrGlyGlyAsnSerGlySerProValPhe 240

RESULT 3
Q99V45
ID   Q99V45     PRELIMINARY;      PRT;   342 AA.
AC   Q99V45;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Serine protease, V8 protease, glutamyl endopeptidase.
GN   SSPA OR SAV1048 OR SA0901.
OS   Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS   Staphylococcus aureus (strain N315).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=158878, 158879;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA   Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA   Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA   Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA   Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA   Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
CC   -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR   EMBL; AP003361; BAB57210.1; -.
DR   EMBL; AP003132; BAB42146.1; -.
DR   MEROPS; S01.269; -.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   InterPro; IPR000126; Ser_proteas_V8.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00839; V8PROTEASE.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00672; V8_HIS; 1.
DR   PROSITE; PS00673; V8_SER; 1.
KW   Hydrolase; Protease; Serine protease; Complete proteome.
SQ   SEQUENCE   342 AA;  36977 MW;  5AEF42DCE01C4B24 CRC64;

Alignment Scores:
Pred. No.:              0.594         Length:        342
Score:                  10.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.40%         Indels:        0
DB:                     16            Gaps:          0

US-10-008-355-1 (1-2139) x Q99V45 (1-342)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    233 ThrGlyGlyAsnSerGlySerProValPhe 242

RESULT 4
```

```
Q04186
ID   Q04186      PRELIMINARY;      PRT;   357 AA.
AC   Q04186;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Glutamic acid specific protease prepropeptide (EC 3.4.21.19).
OS   Staphylococcus aureus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yoshikawa K., Tsuzuki H., Fujiwara T., Nakamura E., Iwamoto H.,
RA   Matsumoto K., Shin M., Yoshida N., Teraoka H.;
RT   "Purification, characterization and gene cloning of a novel glutamic
RT   acid-specific endopeptidase from staphylococcus aureus atcc 12600.";
RL   Biochim. Biophys. Acta 1121:221-228(1991).
DR   EMBL; D00730; BAA00630.1; -.
DR   MEROPS; S01.269; -.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   InterPro; IPR000126; Ser_proteas_V8.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00839; V8PROTEASE.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00672; V8_HIS; 1.
DR   PROSITE; PS00673; V8_SER; 1.
KW   Hydrolase; Protease; Serine protease.
FT   CHAIN        69    357       POTENTIAL.
SQ   SEQUENCE   357 AA;  38651 MW;  58AA9A4E371E2577 CRC64;

Alignment Scores:
Pred. No.:              0.591         Length:        357
Score:                  10.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.40%         Indels:        0
DB:                     2             Gaps:          0

US-10-008-355-1 (1-2139) x Q04186 (1-357)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    233 ThrGlyGlyAsnSerGlySerProValPhe 242

RESULT 5
Q9RU39
ID   Q9RU39      PRELIMINARY;      PRT;   569 AA.
AC   Q9RU39;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   TWO-component sensor, putative.
GN   DR1556.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
CC   -!- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
DR   EMBL; AE001999; AAF11119.1; -.
DR   TIGR; DR1556; -.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR004358; Bact_sens_pr_C.
DR   InterPro; IPR003018; GAF.
DR   InterPro; IPR004359; HIS_KIN_sig.
DR   Pfam; PF02518; HATPase_c; 1.
DR   PRINTS; PR00344; BCTRLSENSOR.
DR   SMART; SM00065; GAF; 1.
DR   SMART; SM00387; HATPase_c; 1.
KW   Kinase; Phosphorylation; Sensory transduction; Transferase;
KW   Complete proteome.
SQ   SEQUENCE   569 AA;  59182 MW;  1A6CA4F88D96A940 CRC64;

Alignment Scores:
Pred. No.:              0.559         Length:        569
Score:                  10.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.40%         Indels:        0
DB:                     16            Gaps:          0

US-10-008-355-1 (1-2139) x Q9RU39 (1-569)

Qy     25 CTTCTCGGAGCAGCCCTGCTGTTGGGTGCT 54
          ||||||||||||||||||||||||||||||
Db     63 LeuLeuGlyAlaAlaLeuLeuLeuGlyAla 72

RESULT 6
Q9AJX0
ID   Q9AJX0      PRELIMINARY;      PRT;   217 AA.
AC   Q9AJX0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Extracellular serine proteinase precursor (Fragment).
GN   ESP.
OS   Staphylococcus epidermidis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=1282;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6746;
RA   Dubin G.;
RT   "Staphylococcus epidermidis extracellular serine proteinase.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ305145; CAC27157.1; -.
DR   HSSP; P09331; 1EXF.
DR   MEROPS; S01.269; -.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   InterPro; IPR000126; Ser_proteas_V8.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00839; V8PROTEASE.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00673; V8_SER; 1.
KW   Hydrolase; Serine protease; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1      1       POTENTIAL.
FT   CHAIN         2    217       EXTRACELLULAR SERINE PROTEINASE.
SQ   SEQUENCE   217 AA;  23667 MW;  FB9B886D453B8BB7 CRC64;

Alignment Scores:
Pred. No.:              7.39          Length:        217
Score:                  9.00          Matches:       9
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.26%         Indels:        0
DB:                     2             Gaps:          0

US-10-008-355-1 (1-2139) x Q9AJX0 (1-217)

Qy   1933 GGCGGTAACTCCGGTAGCCCCGTATTC 1959
```

```
              |||||||||||||||||||||||||||
Db     167 GlyGlyAsnSerGlySerProValPhe 175

RESULT 7
Q9RWF5
ID   Q9RWF5      PRELIMINARY;      PRT;   266 AA.
AC   Q9RWF5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein DR0714.
GN   DR0714.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001928; AAF10302.1; -.
DR   TIGR; DR0714; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   266 AA;  28225 MW;  F7C0ED7F231ADD3C CRC64;

Alignment Scores:
Pred. No.:              7.21          Length:        266
Score:                  9.00          Matches:       9
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.26%         Indels:        0
DB:                     16            Gaps:          0

US-10-008-355-1 (1-2139) x Q9RWF5 (1-266)

Qy      25 CTTCTCGGAGCAGCCCTGCTGTTGGGT 51
              |||||||||||||||||||||||||||
Db     108 LeuLeuGlyAlaAlaLeuLeuLeuGly 116

RESULT 8
Q929Q0
ID   Q929Q0      PRELIMINARY;      PRT;   452 AA.
AC   Q929Q0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein lin2224.
GN   LIN2224.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596171; CAC97453.1; -.
DR   ListiList; LIN02224; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   452 AA;  48806 MW;  0CC0B8BBF765745A CRC64;

Alignment Scores:
Pred. No.:              6.78          Length:        452
Score:                  9.00          Matches:       9
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.26%         Indels:        0
DB:                     16            Gaps:          0

US-10-008-355-1 (1-2139) x Q929Q0 (1-452)

Qy    1555 GTAGAGCTTTCCAAGAGCGTAATAGCT 1581
              |||||||||||||||||||||||||||
Db     142 ValGluLeuSerLysSerValIleAla 150

RESULT 9
Q8YNU9
ID   Q8YNU9      PRELIMINARY;      PRT;   483 AA.
AC   Q8YNU9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein All4462.
GN   ALL4462.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
DR   EMBL; AP003596; BAB76161.1; -.
DR   InterPro; IPR004896; PUCC.
DR   Pfam; PF03209; PUCC; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   483 AA;  51883 MW;  17F82B83C39E7808 CRC64;

Alignment Scores:
Pred. No.:              6.72          Length:        483
Score:                  9.00          Matches:       9
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.26%         Indels:        0
DB:                     16            Gaps:          0

US-10-008-355-1 (1-2139) x Q8YNU9 (1-483)

Qy      34 GCAGCCCTGCTGTTGGGTGCTTCAGGG 60
              |||||||||||||||||||||||||||
Db     353 AlaAlaLeuLeuLeuGlyAlaSerGly 361

RESULT 10
Q9S282
ID   Q9S282      PRELIMINARY;      PRT;   684 AA.
AC   Q9S282;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative integral membrane protein.
GN   SCO1807 OR SCI28.01.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K., Harris D.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL096844; CAB50875.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01582; KV33CHANNEL.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   684 AA;  66534 MW;  F9FC89B471B99752 CRC64;

Alignment Scores:
Pred. No.:               6.45         Length:        684
Score:                   9.00         Matches:       9
Percent Similarity:      100.00%      Conservative:  0
Best Local Similarity:   100.00%      Mismatches:    0
Query Match:             1.26%        Indels:        0
DB:                      16           Gaps:          0

US-10-008-355-1 (1-2139) x Q9S282 (1-684)

Qy      28 CTCGGAGCAGCCCTGCTGTTGGGTGCT 54
           |||||||||||||||||||||||||||
Db     442 LeuGlyAlaAlaLeuLeuLeuGlyAla 450

RESULT 11
Q91WB5
ID   Q91WB5      PRELIMINARY;      PRT;    92 AA.
AC   Q91WB5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical 10.1 kDa protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SALIVARY GLAND;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC016132; AAH16132.1; -.
DR   InterPro; IPR002412; FEL_allergen.
DR   InterPro; IPR000329; Uteroglobin.
DR   Pfam; PF01099; Uteroglobin; 1.
DR   ProDom; PD019935; FEL_allergen; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   92 AA;  10080 MW;  F1A7557E0F5568BB CRC64;

Alignment Scores:
Pred. No.:               96.5         Length:        92
Score:                   8.00         Matches:       8
Percent Similarity:      100.00%      Conservative:  0
Best Local Similarity:   100.00%      Mismatches:    0
Query Match:             1.12%        Indels:        0
DB:                      11           Gaps:          0

US-10-008-355-1 (1-2139) x Q91WB5 (1-92)

Qy      25 CTTCTCGGAGCAGCCCTGCTGTTG 48
           ||||||||||||||||||||||||
Db       9 LeuLeuGlyAlaAlaLeuLeuLeu 16

RESULT 12
Q9X868
ID   Q9X868      PRELIMINARY;      PRT;   102 AA.
AC   Q9X868;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SCO3512.
GN   SCO3512 OR SCE134.13.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL049661; CAB41211.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   102 AA;  10274 MW;  848AD9BBE5308398 CRC64;

Alignment Scores:
Pred. No.:               95.3         Length:        102
Score:                   8.00         Matches:       8
Percent Similarity:      100.00%      Conservative:  0
Best Local Similarity:   100.00%      Mismatches:    0
Query Match:             1.12%        Indels:        0
DB:                      16           Gaps:          0

US-10-008-355-1 (1-2139) x Q9X868 (1-102)

Qy      31 GGAGCAGCCCTGCTGTTGGGTGCT 54
           ||||||||||||||||||||||||
Db      79 GlyAlaAlaLeuLeuLeuGlyAla 86
```

```
RESULT 13
Q9SZ21
ID   Q9SZ21       PRELIMINARY;      PRT;   106 AA.
AC   Q9SZ21;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Hypothetical 12.2 kDa protein.
GN   F10M23.150 OR AT4G26810.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Lecharny A., Chefdor F., Krivitzky M., Kreis M.,
RA   Hoheisel J., Mewes H.W., Mayer K.F.X., Schueller C.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Lecharny A., Chefdor F., Krivitzky M., Kreis M., Mewes H.W.,
RA   Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035440; CAB36527.1; -.
DR   EMBL; AL161565; CAB79536.1; -.
DR   InterPro; IPR003121; SWIB.
DR   Pfam; PF02201; SWIB; 1.
DR   SMART; SM00151; SWIB; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   106 AA;  12209 MW;  8058B0ECA90F4CE1 CRC64;

Alignment Scores:
Pred. No.:              94.9          Length:         106
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            1.12%         Indels:         0
DB:                     10            Gaps:           0

US-10-008-355-1 (1-2139) x Q9SZ21 (1-106)

Qy     10 AAATTAAAAAGTATTCTTCTCGGA 33
          ||||||||||||||||||||||||
Db     71 LysLeuLysSerIleLeuLeuGly 78

RESULT 14
O28315
ID   O28315       PRELIMINARY;      PRT;   162 AA.
AC   O28315;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein AF1964.
GN   AF1964.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
DR   EMBL; AE000967; AAB89289.1; -.
DR   TIGR; AF1964; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   162 AA;  18024 MW;  46404F9FC3EE74AA CRC64;

Alignment Scores:
Pred. No.:              90.3          Length:         162
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            1.13%         Indels:         0
DB:                     17            Gaps:           0

US-10-008-355-1 (1-2139) x O28315 (1-162)

Qy   1401 CTTGAATATATCGGGGAGCTTGTC 1378
          ||||||||||||||||||||||||
Db    142 LeuGluTyrIleGlyGluLeuVal 149

RESULT 15
Q8ZVA1
ID   Q8ZVA1       PRELIMINARY;      PRT;   177 AA.
AC   Q8ZVA1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Transport protein part 1, authentic frameshift.
GN   PAE2385A.
OS   Pyrobaculum aerophilum.
OC   Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC   Thermoproteaceae; Pyrobaculum.
OX   NCBI_TaxID=13773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX   PubMed=11792869;
RA   Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA   Miller J.H.;
RT   "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT   aerophilum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR   EMBL; AE009873; AAL64155.1; -.
KW   Complete proteome.
SQ   SEQUENCE   177 AA;  19519 MW;  FD2BB9B026DEC30F CRC64;

Alignment Scores:
Pred. No.:              89.3          Length:         177
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            1.12%         Indels:         0
DB:                     17            Gaps:           0

US-10-008-355-1 (1-2139) x Q8ZVA1 (1-177)

Qy   1039 GACGTGATAGGTCGTAAGCGTGCC 1062
          ||||||||||||||||||||||||
Db     69 AspValIleGlyArgLysArgAla 76
```

Search completed: May 16, 2003, 13:13:33
Job time : 177 secs

D
P


```
OM nucleic - protein search, using frame_plus_n2p model

Run on:         May 16, 2003, 12:39:16 ; Search time 36.5 Seconds
                                          (without alignments)
                                          4861.253 Million cell updates/sec

Title:          US-10-008-355-1
Perfect score:  712
Sequence:       1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:  OLIGO
                Xgapop 60.0 , Xgapext 60.0
                Ygapop 60.0 , Ygapext 60.0
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       112892 seqs, 41476328 residues

Word size:      1

Total number of hits satisfying chosen parameters:      225614

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110425_9723/app_query.fasta_1.2311
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=olin2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_47_@runat_16052003_110425_9723 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :      SwissProt_40:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	
	1	10	1.4	336	1	STSP_STAAU	P04188	staphylococ
	2	9	1.3	976	1	EPA1_HUMAN	P21709	homo sapien
	3	9	1.3	1116	1	YK54_AQUAE	O67838	aquifex aeo
	4	8	1.1	116	1	CART_HUMAN	Q16568	homo sapien
	5	8	1.1	129	1	CART_MOUSE	P56388	mus musculu
	6	8	1.1	214	1	TPIS_HALN1	Q9hqs4	halobacteri
c	7	8	1.1	226	1	ATP6_RAT	P05504	rattus norv
	8	8	1.1	256	1	YDH1_HSVSC	P22575	herpesvirus
	9	8	1.1	279	1	Y939_METKA	P58851	methanopyru
	10	8	1.1	349	1	YNJ6_CAEEL	P34550	caenorhabdi
	11	8	1.1	382	1	METK_SCHPO	O60198	schizosacch
	12	8	1.1	392	1	CHSY_PERAE	Q9zu06	persea amer
	13	8	1.1	393	1	METK_ASCIM	P50304	ascobolus i
	14	8	1.1	537	1	MPPA_ECOLI	P77348	escherichia
c	15	8	1.1	570	1	FBP3_STRPU	P49013	strongyloce
c	16	8	1.1	586	1	RUBA_ARATH	P21238	arabidopsis
	17	8	1.1	637	1	ILV2_BRANA	P14874	brassica na
c	18	8	1.1	757	1	MUS2_THEMA	Q9x105	thermotoga
	19	8	1.1	876	1	AREA_EMENI	P17429	emericella
c	20	8	1.1	900	1	SYA_AERPE	Q9y9x3	aeropyrum p
	21	8	1.1	969	1	MNR2_YEAST	P35724	saccharomyc
c	22	7	1.0	32	1	Y16O_BPT4	P39247	bacteriopha
c	23	7	1.0	42	1	RLA2_WHEAT	P05390	triticum ae
	24	7	1.0	68	1	RL29_STRPY	Q9a1w6	streptococc
c	25	7	1.0	94	1	GATC_CAMJE	Q9pia5	campylobact
	26	7	1.0	113	1	PSAK_CHLRE	P14225	chlamydomon
	27	7	1.0	116	1	Y1B6_MYCLE	O69519	mycobacteri
c	28	7	1.0	119	1	BCP1_BRACM	P41507	brassica ca
c	29	7	1.0	123	1	RS13_MYCGE	P47421	mycoplasma
c	30	7	1.0	124	1	URE2_BACSU	P71035	bacillus su
c	31	7	1.0	127	1	Y55B_MYCPN	P75047	mycoplasma
	32	7	1.0	129	1	CART_RAT	P49192	rattus norv
c	33	7	1.0	136	1	SCPB_APLSP	P09892	aplysia sp.
c	34	7	1.0	139	1	NUSA_HALMO	P15738	halococcus
c	35	7	1.0	139	1	NUSA_HALN1	P15739	halobacteri
c	36	7	1.0	142	1	H2B2_LYTPI	P06146	lytechinus
c	37	7	1.0	142	1	U426_HSVMG	Q05104	marek's dis
c	38	7	1.0	143	1	NUSA_SULAC	P11523	sulfolobus
	39	7	1.0	147	1	HFAA_CAUCR	P27342	caulobacter
c	40	7	1.0	167	1	LSPA_CHLTR	O84413	chlamydia t
c	41	7	1.0	167	1	NU6M_BRALA	O79423	branchiosto
c	42	7	1.0	169	1	YSFO_STRCO	Q9zbg3	streptomyce
c	43	7	1.0	173	1	YOJ3_PSEAE	Q9hxu7	pseudomonas
c	44	7	1.0	177	1	Y189_STRGR	Q9khe2	streptomyce
	45	7	1.0	177	1	YH21_AZOCH	Q43924	azotobacter

ALIGNMENTS

```
RESULT 1
STSP_STAAU
ID   STSP_STAAU     STANDARD;      PRT;   336 AA.
AC   P04188;
DT   20-MAR-1987 (Rel. 04, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Glutamyl endopeptidase precursor (EC 3.4.21.19) (Staphylococcal serine
DE   proteinase) (V8 proteinase) (Endoproteinase Glu-C).
GN   SSPA.
OS   Staphylococcus aureus.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=V8;
RX   MEDLINE=87316953; PubMed=3306605;
RA   Carmona C., Gray G.L.;
RT   "Nucleotide sequence of the serine protease gene of Staphylococcus
RT   aureus, strain V8.";
RL   Nucleic Acids Res. 15:6757-6757(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20569178; PubMed=11119502;
RA   Rice K., Peralta R., Bast D., de Azavedo J., McGavin M.J.;
RT   "Description of staphylococcus serine protease (ssp) operon in
RT   Staphylococcus aureus and nonpolar inactivation of sspA-encoded serine
RT   protease.";
RL   Infect. Immun. 69:159-169(2001).
RN   [3]
RP   SEQUENCE OF 69-280.
RC   STRAIN=V8;
RX   MEDLINE=78212487; PubMed=96922;
RA   Drapeau G.R.;
RT   "The primary structure of staphylococcal protease.";
RL   Can. J. Biochem. 56:534-544(1978).
CC   -!- FUNCTION: PREFERENTIALLY CLEAVES PEPTIDE BONDS ON THE CARBOXYL-
CC       TERMINAL SIDE OF ASPARTATE AND GLUTAMATE.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Asp-|-Xaa, Glu-|-Xaa.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/P/STAP.html".
```

```
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Y00356; CAA68434.1; -.            P04188
DR   EMBL; AF309515; AAG45843.1; -.
DR   PIR; A26812; PRSASK.
DR   MEROPS; S01.269; -.
DR   InterPro; IPR000126; Ser_proteas_V8.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00839; V8PROTEASE.
DR   PROSITE; PS00672; V8_HIS; 1.
DR   PROSITE; PS00673; V8_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal.
FT   SIGNAL        1     29       POTENTIAL.
FT   PROPEP       30     68
FT   CHAIN        69    336       GLUTAMYL ENDOPEPTIDASE.
FT   ACT_SITE    119    119       CHARGE RELAY SYSTEM.
FT   ACT_SITE    161    161       CHARGE RELAY SYSTEM.
FT   ACT_SITE    237    237       CHARGE RELAY SYSTEM.
FT   CONFLICT    109    109       MISSING (IN REF. 3).
FT   CONFLICT    125    125       MISSING (IN REF. 3).
FT   CONFLICT    145    145       N -> D (IN REF. 3).
FT   CONFLICT    193    193       V -> T (IN REF. 3).
FT   CONFLICT    229    229       D -> N (IN REF. 3).
FT   CONFLICT    259    261       EFN -> QFD (IN REF. 3).
FT   CONFLICT    268    270       ENV -> NEVN (IN REF. 3).
SQ   SEQUENCE   336 AA;  36326 MW;  8B138D0C7996AA3E CRC64;

Alignment Scores:
Pred. No.:               0.228        Length:        336
Score:                   10.00        Matches:       10
Percent Similarity:      100.00%      Conservative:  0
Best Local Similarity:   100.00%      Mismatches:    0
Query Match:             1.40%        Indels:        0
DB:                      1            Gaps:          0

US-10-008-355-1 (1-2139) x STSP_STAAU (1-336)

Qy   1930 ACGGGCGGTAACTCCGGTAGCCCCGTATTC 1959
          ||||||||||||||||||||||||||||||
Db    233 ThrGlyGlyAsnSerGlySerProValPhe 242

RESULT 2
EPA1_HUMAN
ID   EPA1_HUMAN     STANDARD;      PRT;   976 AA.
AC   P21709; Q15405;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ephrin type-A receptor 1 precursor (EC 2.7.1.112) (Tyrosine-protein
DE   kinase receptor EPH).
GN   EPHA1 OR EPHT1 OR EPHT OR EPH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88070650; PubMed=2825356;
RA   Hirai H., Maru Y., Hagiwara K., Nishida J., Takaku F.;
RT   "A novel putative tyrosine kinase receptor encoded by the eph gene.";
RL   Science 238:1717-1720(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99299440; PubMed=10369740;
RA   Owshalimpur D., Kelley M.J.;
RT   "Genomic structure of the EPHA1 receptor tyrosine kinase gene.";
RL   Mol. Cell. Probes 13:169-173(1999).
RN   [3]
RP   SEQUENCE OF 286-976 FROM N.A.
RC   TISSUE=Placenta;
RA   Tuzi N.L.;
RL   Submitted (MAY-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RECEPTOR FOR MEMBERS OF THE EPHRIN-A FAMILY. BINDS WITH
CC       A LOW AFFINITY TO EPHRIN-A1.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC   -!- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. EPHRIN
CC       RECEPTOR SUBFAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M18391; AAA36747.1; ALT_SEQ.
DR   EMBL; AF101171; AAD43440.1; -.
DR   EMBL; AF101165; AAD43440.1; JOINED.
DR   EMBL; AF101166; AAD43440.1; JOINED.
DR   EMBL; AF101167; AAD43440.1; JOINED.
DR   EMBL; AF101168; AAD43440.1; JOINED.
DR   EMBL; AF101169; AAD43440.1; JOINED.
DR   EMBL; AF101170; AAD43440.1; JOINED.
DR   EMBL; Z27409; CAA81796.1; -.
DR   PIR; A34076; A34076.
DR   HSSP; P00523; 2PTK.
DR   Genew; HGNC:3385; EPHA1.
DR   MIM; 179610; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001090; Ephrin_receptor.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR001660; SAM.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR001426; YKase_receptorV.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00536; SAM; 1.
DR   Pfam; PF01404; EPH_lbd; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   ProDom; PD001495; Ephrin_receptor; 1.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00454; SAM; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00790; RECEPTOR_TYR_KIN_V_1; 1.
DR   PROSITE; PS00791; RECEPTOR_TYR_KIN_V_2; 1.
DR   PROSITE; PS50105; SAM_DOMAIN; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Receptor; Transmembrane; Glycoprotein; Signal; Repeat.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    976       EPHRIN TYPE-A RECEPTOR 1.
FT   DOMAIN       24    547       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    548    568       POTENTIAL.
FT   DOMAIN      569    976       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      191    329       CYS-RICH.
FT   DOMAIN      330    442       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      443    537       FIBRONECTIN TYPE-III 2.
```

```
FT   DOMAIN      624    884       PROTEIN KINASE.
FT   DOMAIN      913    976       SAM.
FT   SITE        974    976       PDZ-BINDING MOTIF (POTENTIAL).
FT   NP_BIND     630    638       ATP (BY SIMILARITY).
FT   BINDING     656    656       ATP (BY SIMILARITY).
FT   ACT_SITE    749    749       BY SIMILARITY.
FT   MOD_RES     599    599       PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT   MOD_RES     605    605       PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT   MOD_RES     781    781       PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT   MOD_RES     930    930       PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT   CARBOHYD    338    338       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    398    398       A -> G (IN REF. 1).
FT   CONFLICT    581    616       QRDRATDVDREDKLWLKPYVDLQAYEDPAQGALDFT ->
FT                                HVTAPPMWIERTSCAEALCGTSRHTRTLHREPWTLPGGWSN
FT                                FPS (IN REF. 1).
SQ   SEQUENCE   976 AA;  108080 MW;  50512CCF13F03808 CRC64;

Alignment Scores:
Pred. No.:              2.1          Length:        976
Score:                  9.00         Matches:       9
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            1.26%        Indels:        0
DB:                     1            Gaps:          0

US-10-008-355-1 (1-2139) x EPA1_HUMAN (1-976)

Qy      25 CTTCTCGGAGCAGCCCTGCTGTTGGGT 51
           |||||||||||||||||||||||||||
Db     556 LeuLeuGlyAlaAlaLeuLeuLeuGly 564

RESULT 3
YK54_AQUAE
ID   YK54_AQUAE     STANDARD;      PRT;  1116 AA.
AC   O67838;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein AQ_2054.
GN   AQ_2054.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC   Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AE000770; AAC07805.1; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM      3     20       POTENTIAL.
SQ   SEQUENCE   1116 AA;  127794 MW;  423E5FCC6C02C89C CRC64;

Alignment Scores:
Pred. No.:              2.06         Length:        1116
Score:                  9.00         Matches:       9
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            1.26%        Indels:        0
DB:                     1            Gaps:          0

US-10-008-355-1 (1-2139) x YK54_AQUAE (1-1116)

Qy    1405 GTAATCGACAAGAAATTCAAAGGCGAC 1431
           |||||||||||||||||||||||||||
Db     329 ValIleAspLysLysPheLysGlyAsp 337

RESULT 4
CART_HUMAN
ID   CART_HUMAN     STANDARD;      PRT;   116 AA.
AC   Q16568;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cocaine- and amphetamine-regulated transcript protein precursor
DE   [Contains: CART(1-39); CART(32-89)].
GN   CART.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96194810; PubMed=8647455;
RA   Douglass J.O., Daoud S.;
RT   "Characterization of the human cDNA and genomic DNA encoding CART: a
RT   cocaine- and amphetamine-regulated transcript.";
RL   Gene 169:241-245(1996).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=98250573; PubMed=9590691;
RA   Kristensen P., Judge M.E., Thim L., Ribel U., Christjansen K.N.,
RA   Wulff B.S., Clausen J.T., Jensen P.B., Madsen O.D., Vrang N.,
RA   Larsen P.J., Hastrup S.;
RT   "Hypothalamic CART is a new anorectic peptide regulated by leptin.";
RL   Nature 393:72-76(1998).
RN   [3]
RP   STRUCTURE BY NMR OF 75-116.
RX   MEDLINE=21371766; PubMed=11478874;
RA   Ludvigsen S., Thim L., Blom A.M., Wulff B.S.;
RT   "Solution structure of the satiety factor, CART, reveals new
RT   functionality of a well-known fold.";
RL   Biochemistry 40:9082-9088(2001).
RN   [4]
RP   VARIANT THR-66.
RX   MEDLINE=20361434; PubMed=10905499;
RA   Challis B.G., Yeo G.S.H., Farooqi I.S., Luan J., Aminian S.,
RA   Halsall D.J., Keogh J.M., Wareham N.J., O'Rahilly S.;
RT   "The CART gene and human obesity: mutational analysis and population
RT   genetics.";
RL   Diabetes 49:872-875(2000).
RN   [5]
RP   VARIANT PHE-61.
RX   MEDLINE=21413627; PubMed=11522684;
RA   del Giudice E.M., Santoro N., Cirillo G., D'Urso L., Di Toro R.,
RA   Perrone L.;
RT   "Mutational screening of the CART gene in obese children: identifying
RT   a mutation (Leu34Phe) associated with reduced resting energy
RT   expenditure and cosegregating with obesity phenotype in a large
RT   family.";
RL   Diabetes 50:2157-2160(2001).
CC   -!- FUNCTION: Satiety factor closely associated with the actions of
CC       leptin and neuropeptide y; this anorectic peptide inhibits both
CC       normal and starvation-induced feeding and completely blocks the
CC       feeding response induced by neuropeptide Y and regulated by leptin
CC       in the hypothalamus. It promotes neuronal development and survival
CC       in vitro.
CC   -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC   -!- TISSUE SPECIFICITY: Hypothalamus. Found in neurons of the
```

```
CC       ventrolateral part of the arcuate nucleus, in the external zone of
CC       the median eminence, and also found in terminals in the
CC       periventricular part of the paraventricular nucleus.
CC   -!- INDUCTION: By leptin.
CC   -!- DISEASE: A defect in CART is associated with reduced resting
CC       energy expenditure and cosegregates with obesity phenotype.
CC   -!- SIMILARITY: BELONGS TO THE CART FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U16826; AAB08010.1; -.
DR   EMBL; U20325; AAB08011.1; -.
DR   PDB; 1HY9; 29-AUG-01.
DR   MIM; 602606; -.
KW   Neuropeptide; Neurotransmitter; Cleavage on pair of basic residues;
KW   Signal; Disease mutation; Polymorphism; Obesity; 3D-structure.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    116       COCAINE- AND AMPHETAMINE-REGULATED
FT                                TRANSCRIPT PROTEIN.
FT   PEPTIDE      28     66       CART(1-39).
FT   PEPTIDE      69    116       CART(42-89).
FT   DISULFID     82    100
FT   DISULFID     88    108
FT   DISULFID    102    115
FT   VARIANT      61     61       L -> F (IN AN OBESE PATIENT).
FT                                /FTId=VAR_012199.
FT   VARIANT      66     66       S -> T.
FT                                /FTId=VAR_012200.
SQ   SEQUENCE   116 AA;  12829 MW;  FC396CA2C032AA83 CRC64;

Alignment Scores:
Pred. No.:              30.1           Length:          116
Score:                  8.00           Matches:         8
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            1.12%          Indels:          0
DB:                     1              Gaps:            0

US-10-008-355-1 (1-2139) x CART_HUMAN (1-116)

Qy      25 CTTCTCGGAGCAGCCCTGCTGTTG 48
           ||||||||||||||||||||||||
Db      11 LeuLeuGlyAlaAlaLeuLeuLeu 18

RESULT 5
CART_MOUSE
ID   CART_MOUSE     STANDARD;      PRT;   129 AA.
AC   P56388; Q9QXZ8;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cocaine- and amphetamine-regulated transcript protein precursor
DE   [Contains: CART(1-52); CART(55-102)].
GN   CART.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   STRAIN=129/SvJ;
RX   MEDLINE=20080750; PubMed=10612705;
RA   Adams L.D., Gong W., Vechia S.D., Hunter R.G., Kuhar M.J.;
RT   "CART: from gene to function.";
RL   Brain Res. 848:137-140(1999).
CC   -!- FUNCTION: Satiety factor closely associated with the actions of
CC       leptin and neuropeptide y; this anorectic peptide inhibits both
CC       normal and starvation-induced feeding and completely blocks the
CC       feeding response induced by neuropeptide Y and regulated by leptin
CC       in the hypothalamus (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
CC       short form; are produced by alternative splicing.
CC   -!- SIMILARITY: BELONGS TO THE CART FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF148071; AAF24168.1; -.
DR   HSSP; Q16568; 1HY9.
DR   MGD; MGI:1351330; Cart.
KW   Neuropeptide; Neurotransmitter; Cleavage on pair of basic residues;
KW   Alternative splicing; Signal.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    129       COCAINE- AND AMPHETAMINE-REGULATED
FT                                TRANSCRIPT PROTEIN.
FT   PEPTIDE      28     79       CART(1-52) (BY SIMILARITY).
FT   PEPTIDE      82    129       CART(55-102) (BY SIMILARITY).
FT   PEPTIDE      89    129       CART(62-102) (POTENTIAL).
FT   DISULFID     95    113       BY SIMILARITY.
FT   DISULFID    101    121       BY SIMILARITY.
FT   DISULFID    115    128       BY SIMILARITY.
FT   VARSPLIC     54     66       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   129 AA;  14285 MW;  E4CA6CE70BEE6DF2 CRC64;

Alignment Scores:
Pred. No.:              29.7           Length:          129
Score:                  8.00           Matches:         8
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            1.12%          Indels:          0
DB:                     1              Gaps:            0

US-10-008-355-1 (1-2139) x CART_MOUSE (1-129)

Qy      25 CTTCTCGGAGCAGCCCTGCTGTTG 48
           ||||||||||||||||||||||||
Db      11 LeuLeuGlyAlaAlaLeuLeuLeu 18

RESULT 6
TPIS_HALN1
ID   TPIS_HALN1     STANDARD;      PRT;   214 AA.
AC   Q9HQS4;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Triosephosphate isomerase (EC 5.3.1.1) (TIM).
GN   TPIA OR VNG1027G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
```

```
CC   -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycerone
CC       phosphate.
CC   -!- PATHWAY: Plays an important role in several metabolic pathways.
CC   -!- SUBUNIT: Homotetramer (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE005037; AAG19439.1; -.
DR   InterPro; IPR003009; FMN_enzyme.
DR   InterPro; IPR002173; PfkB.
DR   InterPro; IPR000652; Triophos_ismrse.
DR   ProDom; PD001005; Triophos_ismrse; 1.
DR   TIGRFAMs; TIGR00419; tim; 1.
DR   PROSITE; PS00171; TIM; FALSE_NEG.
KW   Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KW   Pentose shunt; Complete proteome.
FT   ACT_SITE     85     85       BY SIMILARITY.
SQ   SEQUENCE   214 AA;  20928 MW;  B6DD20B1DD85D6A4 CRC64;

Alignment Scores:
Pred. No.:              27.7          Length:        214
Score:                  8.00          Matches:       8
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.12%         Indels:        0
DB:                     1             Gaps:          0

US-10-008-355-1 (1-2139) x TPIS_HALN1 (1-214)

Qy      52 GCTTCAGGGGTAGCCAAAGCCGAC 75
           ||||||||||||||||||||||||
Db     194 AlaSerGlyValAlaLysAlaAsp 201

RESULT 7
ATP6_RAT
ID   ATP6_RAT      STANDARD;      PRT;   226 AA.
AC   P05504; Q9T2E9;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   ATP synthase A chain (EC 3.6.3.14) (Protein 6).
GN   MTATP6 OR ATP6.
OS   Rattus norvegicus (Rat).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar;
RX   MEDLINE=89362487; PubMed=2504926;
RA   Gadaleta G., Pepe G., de Candia G., Quagliariello C., Sbisa E.,
RA   Saccone C.;
RT   "The complete nucleotide sequence of the Rattus norvegicus
RT   mitochondrial genome: cryptic signals revealed by comparative
RT   analysis between vertebrates.";
RL   J. Mol. Evol. 28:497-516(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RA   Grosskopf R., Feldmann H.;
RT   "Analysis of a DNA segment from rat liver mitochondria containing the
RT   genes for the cytochrome oxidase subunits I, II, II, ATPase subunit
RT   6, and several tRNA genes.";
RL   Curr. Genet. 4:151-158(1981).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANT ASN-101.
RC   STRAIN=BHE/CDB, and Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=96118462; PubMed=8529844;
RA   Mathews C.E., McGraw R.A., Berdanier C.D.;
RT   "A point mutation in the mitochondrial DNA of diabetes-prone BHE/cdb
RT   rats.";
RL   FASEB J. 9:1638-1642(1995).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.55 ANGSTROMS) OF 29-41.
RX   MEDLINE=21109741; PubMed=11163232;
RA   Speir J.A., Stevens J., Joly E., Butcher G.W., Wilson I.A.;
RT   "Two different, highly exposed, bulged structures for an unusually
RT   long peptide bound to rat MHC class I RT1-A(a).";
RL   Immunity 14:81-92(2001).
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC       H(+)(Out).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J01435; AAD15019.1; -.
DR   EMBL; X14848; CAA32959.1; -.
DR   EMBL; AF115770; AAD22965.1; -.
DR   PIR; S04752; S04752.
DR   PDB; 1ED3; 21-JUN-00.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane;
KW   3D-structure.
FT   VARIANT     101    101       D -> N.
FT   CONFLICT      7      7       A -> P (IN REF. 2).
FT   CONFLICT    129    129       S -> L (IN REF. 2).
FT   CONFLICT    205    205       A -> V (IN REF. 2).
SQ   SEQUENCE   226 AA;  25050 MW;  6074E2CCAC2B586E CRC64;

Alignment Scores:
Pred. No.:              27.4          Length:        226
Score:                  8.00          Matches:       8
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.13%         Indels:        0
DB:                     1             Gaps:          0

US-10-008-355-1 (1-2139) x ATP6_RAT (1-226)

Qy    1517 TTGTCCATGGACTTGAGCATGGCA 1494
           ||||||||||||||||||||||||
Db      98 LeuSerMetAspLeuSerMetAla 105

RESULT 8
YDH1_HSVSC
ID   YDH1_HSVSC     STANDARD;      PRT;   256 AA.
AC   P22575;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 28.7 kDa protein in DHFR 3'region (ORF1).
```

```
OS   Herpesvirus saimiri (subgroup C / strain 488).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=10384;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90266466; PubMed=2161148;
RA   Biesinger B., Trimble J.J., Desrosiers R.C., Fleckenstein B.;
RT   "The divergence between two oncogenic Herpesvirus saimiri strains in
RT   a genomic region related to the transforming phenotype.";
RL   Virology 176:505-514(1990).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M55264; AAA72928.1; -.
DR   PIR; A34770; A34770.
KW   Hypothetical protein.
SQ   SEQUENCE   256 AA;  28662 MW;  E04BA49D27A59D3A CRC64;

Alignment Scores:
Pred. No.:              27          Length:         256
Score:                  8.00        Matches:        8
Percent Similarity:     100.00%     Conservative:   0
Best Local Similarity:  100.00%     Mismatches:     0
Query Match:            1.12%       Indels:         0
DB:                     1           Gaps:           0

US-10-008-355-1 (1-2139) x YDH1_HSVSC (1-256)

Qy     548 ACAACGAATACTTCCTCATCGTCT 571
           ||||||||||||||||||||||||
Db      33 ThrThrAsnThrSerSerSerSer 40

RESULT 9
Y939_METKA
ID   Y939_METKA     STANDARD;      PRT;   279 AA.
AC   P58851;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein MK0939.
GN   MK0939.
OS   Methanopyrus kandleri.
OC   Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC   Methanopyrus.
OX   NCBI_TaxID=2320;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AV19 / DSM 6324 / JCM 9639;
RX   MEDLINE=21927647; PubMed=11930014;
RA   Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA   Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA   Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA   Malykh A.G., Koonin E.V., Kozyavkin S.A.;
RT   "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT   and monophyly of archaeal methanogens.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC   -!- SIMILARITY: BELONGS TO THE UPF0204 FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AE010383; AAM02152.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   279 AA;  31379 MW;  49474EC38B012D85 CRC64;

Alignment Scores:
Pred. No.:              26.7        Length:         279
Score:                  8.00        Matches:        8
Percent Similarity:     100.00%     Conservative:   0
Best Local Similarity:  100.00%     Mismatches:     0
Query Match:            1.12%       Indels:         0
DB:                     1           Gaps:           0

US-10-008-355-1 (1-2139) x Y939_METKA (1-279)

Qy     301 GTGGATCACGACTATCTGCGCGAT 324
           ||||||||||||||||||||||||
Db     224 ValAspHisAspTyrLeuArgAsp 231

RESULT 10
YNJ6_CAEEL
ID   YNJ6_CAEEL     STANDARD;      PRT;   349 AA.
AC   P34550;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 38.0 kDa protein R10E11.6 in chromosome III.
GN   R10E11.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: CONTAINS 1 EH DOMAIN.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z29095; CAA82351.1; -.
DR   PIR; S40718; S40718.
DR   WormPep; R10E11.6; CE00308.
DR   InterPro; IPR000261; EPS15_repeat.
DR   PROSITE; PS50031; EH; 1.
KW   Hypothetical protein.
FT   DOMAIN       43    148       EH.
SQ   SEQUENCE   349 AA;  37984 MW;  862F7D9D3E3E2E2F CRC64;

Alignment Scores:
Pred. No.:              25.8        Length:         349
Score:                  8.00        Matches:        8
Percent Similarity:     100.00%     Conservative:   0
Best Local Similarity:  100.00%     Mismatches:     0
```

```
Query Match:            1.12%         Indels:        0
DB:                     1             Gaps:          0

US-10-008-355-1 (1-2139) x YNJ6_CAEEL (1-349)

Qy     545 CCAACAACGAATACTTCCTCATCG 568
           ||||||||||||||||||||||||
Db     150 ProThrThrAsnThrSerSerSer 157

RESULT 11
METK_SCHPO
ID   METK_SCHPO     STANDARD;      PRT;   382 AA.
AC   O60198;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine
DE   adenosyltransferase) (AdoMet synthetase).
GN   SAM1 OR SPBC14F5.05C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hilti N.;
RT   "sam1, a new gene coding for S-adenosylmethionine synthetase in S.
RT   pombe.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM
CC       METHIONINE AND ATP (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +
CC       diphosphate + S-adenosyl-L-methionine.
CC   -!- PATHWAY: Activated methyl cycle.
CC   -!- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AJ001705; CAA04941.1; -.
DR   EMBL; AL023780; CAA19323.1; -.
DR   HSSP; P04384; 1MXB.
DR   InterPro; IPR002133; S-AdoMet_synt.
DR   Pfam; PF00438; S-AdoMet_synt; 1.
DR   Pfam; PF02772; S-AdoMet_syntD2; 1.
DR   Pfam; PF02773; S-AdoMet_syntD3; 1.
DR   TIGRFAMs; TIGR01034; metK; 1.
DR   PROSITE; PS00376; ADOMET_SYNTHETASE_1; 1.
DR   PROSITE; PS00377; ADOMET_SYNTHETASE_2; 1.
KW   Transferase; One-carbon metabolism; ATP-binding.
FT   NP_BIND      118    123       ATP (POTENTIAL).
FT   BINDING      146    146       ATP (POTENTIAL).
SQ   SEQUENCE   382 AA;  41831 MW;  9970A9D1195C5738 CRC64;

Alignment Scores:
Pred. No.:              25.5          Length:        382
Score:                  8.00          Matches:       8
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            1.12%         Indels:        0
DB:                     1             Gaps:          0

US-10-008-355-1 (1-2139) x METK_SCHPO (1-382)

Qy    1755 CGGTGCCTGGTACAACTATCATAC 1778
           ||||||||||||||||||||||||
Db     300 ArgCysLeuValGlnLeuSerTyr 307

RESULT 12
CHSY_PERAE
ID   CHSY_PERAE     STANDARD;      PRT;   392 AA.
AC   Q9ZU06;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Chalcone synthase (EC 2.3.1.74) (Naregenin-chalcone synthase).
GN   CHS.
OS   Persea americana (Avocado).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX   NCBI_TaxID=3435;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ardi R., Kobiler I., Jacoby B., Keen N.T., Prusky D.;
RT   "Involvement of epicatechin biosynthesis in the activation of the
RT   mechanism of resistance of avocado fruits to Colletotrichum
RT   gloeosporioides.";
RL   Physiol. Mol. Plant Pathol. 53:269-286(1998).
CC   -!- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-
CC       TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE).
CC       WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC       NARINGENIN.
CC   -!- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC       naringenin-chalcone + 3 CO(2).
CC   -!- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC       FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC       OF WHICH ARE BRIGHTLY COLORED.
CC   -!- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AD001672; AAC98143.1; -.
DR   InterPro; IPR001099; N-C_synthase.
DR   Pfam; PF00195; Chal_stil_synt; 1.
```

```
DR   Pfam; PF02797; Chal_stil_syntC; 1.
DR   ProDom; PD000453; N-C_synthase; 1.
DR   PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW   Flavonoid biosynthesis; Transferase; Acyltransferase.
FT   ACT_SITE    165    165       BY SIMILARITY.
SQ   SEQUENCE   392 AA;  42480 MW;  B08D286FB80BB42E CRC64;

Alignment Scores:
Pred. No.:              25.4          Length:         392
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            1.12%         Indels:         0
DB:                     1             Gaps:           0

US-10-008-355-1 (1-2139) x CHSY_PERAE (1-392)

Qy   1223 GTTTTGCACAGTTTGCCAACGCAT 1246
          ||||||||||||||||||||||||
Db    385 ValLeuHisSerLeuProThrHis 392

RESULT 13
METK_ASCIM
ID   METK_ASCIM     STANDARD;      PRT;   393 AA.
AC   P50304;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine
DE   adenosyltransferase) (AdoMet synthetase).
OS   Ascobolus immersus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Pezizomycetes;
OC   Pezizales; Ascobolaceae; Ascobolus.
OX   NCBI_TaxID=5191;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RN42;
RX   MEDLINE=96200878; PubMed=8621082;
RA   Mautino M.R., Goyon C., Rosa A.L.;
RT   "Cloning and sequence of the Ascobolus immersus S-adenosyl-L-
RT   methionine synthetase-encoding gene.";
RL   Gene 170:155-156(1996).
CC   -!- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM
CC       METHIONINE AND ATP.
CC   -!- CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +
CC       diphosphate + S-adenosyl-L-methionine.
CC   -!- PATHWAY: Activated methyl cycle.
CC   -!- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U21548; AAB03805.1; -.
DR   HSSP; P04384; 1MXB.
DR   InterPro; IPR002133; S-AdoMet_synt.
DR   Pfam; PF00438; S-AdoMet_synt; 1.
DR   Pfam; PF02772; S-AdoMet_syntD2; 1.
DR   Pfam; PF02773; S-AdoMet_syntD3; 1.
DR   TIGRFAMs; TIGR01034; metK; 1.
DR   PROSITE; PS00376; ADOMET_SYNTHETASE_1; 1.
DR   PROSITE; PS00377; ADOMET_SYNTHETASE_2; 1.
KW   Transferase; One-carbon metabolism; ATP-binding.
FT   NP_BIND      127    132       ATP (POTENTIAL).
FT   BINDING      155    155       ATP (POTENTIAL).
SQ   SEQUENCE   393 AA;  43010 MW;  88FB2F2F14B751C3 CRC64;

Alignment Scores:
Pred. No.:              25.4          Length:         393
Score:                  8.00          Matches:        8
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            1.12%         Indels:         0
DB:                     1             Gaps:           0

US-10-008-355-1 (1-2139) x METK_ASCIM (1-393)

Qy   1755 CGGTGCCTGGTACAACTATCATAC 1778
          ||||||||||||||||||||||||
Db    309 ArgCysLeuValGlnLeuSerTyr 316

RESULT 14
MPPA_ECOLI
ID   MPPA_ECOLI     STANDARD;      PRT;   537 AA.
AC   P77348;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Periplasmic murein peptide-binding protein precursor.
GN   MPPA OR B1329.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-32.
RC   STRAIN=K12 / AT980;
RX   MEDLINE=98155149; PubMed=9495761;
RA   Park J.T., Raychaudhuri D., Li H., Normark S., Mengin-Lecreulx D.;
RT   "MppA, a periplasmic binding protein essential for import of the
RT   bacterial cell wall peptide L-alanyl-gamma-D-glutamyl-meso-
RT   diaminopimelate.";
RL   J. Bacteriol. 180:1215-1223(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
CC   -!- FUNCTION: ESSENTIAL FOR THE UPTAKE OF THE MUREIN PEPTIDE L-ALANYL-
CC       GAMMA-D-GLUTAMYL-MESO-DIAMINOPIMELATE. ALSO TRANSPORTS SOME ALPHA-
CC       LINKED PEPTIDES SUCH AS PRO-PHE-LYS WITH LOW AFFINITY. THE
CC       TRANSPORT IS EFFECTED BY THE OLIGOPEPTIDE PERMEASE SYSTEM.
CC   -!- SUBCELLULAR LOCATION: Periplasmic.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC       PROTEIN FAMILY 5.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; U88242; AAC38216.1; -.
DR   EMBL; AE000231; AAC74411.1; ALT_INIT.
DR   EMBL; D90772; BAA14932.1; ALT_INIT.
DR   EMBL; D90771; BAA14922.1; ALT_INIT.
DR   HSSP; P06202; 1JEV.
DR   SWISS-2DPAGE; P77348; COLI.
DR   EcoGene; EG13376; mppA.
DR   InterPro; IPR000914; SBP_bac_5.
DR   Pfam; PF00496; SBP_bac_5; 1.
DR   PROSITE; PS01040; SBP_BACTERIAL_5; 1.
KW   Peptide transport; Transport; Periplasmic; Signal; Complete proteome.
FT   SIGNAL        1     22
FT   CHAIN        23    537       PERIPLASMIC MUREIN PEPTIDE-BINDING
FT                                PROTEIN.
SQ   SEQUENCE   537 AA;  59900 MW;  C6A17656836DC3AC CRC64;

Alignment Scores:
Pred. No.:              24.4           Length:        537
Score:                  8.00           Matches:       8
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.12%          Indels:        0
DB:                     1              Gaps:          0

US-10-008-355-1 (1-2139) x MPPA_ECOLI (1-537)

Qy   1892 GTCAGCTCCATATCGCTTTCCTAT 1915
          ||||||||||||||||||||||||
Db     14 ValSerSerIleSerLeuSerTyr 21

RESULT 15
FBP3_STRPU
ID   FBP3_STRPU     STANDARD;      PRT;   570 AA.
AC   P49013;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Fibropellin C precursor (Epidermal growth factor-related protein 3)
DE   (EGF III) (Fibropellin III).
GN   EGF3.
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Gastrula;
RX   MEDLINE=93273088; PubMed=8500658;
RA   Bisgrove B.W., Raff R.A.;
RT   "The SpEGF III gene encodes a member of the fibropellins: EGF repeat-
RT   containing proteins that form the apical lamina of the sea urchin
RT   embryo.";
RL   Dev. Biol. 157:526-538(1993).
CC   -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC       MATRIX.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN UNFERTILIZED EGGS AND DURING
CC       EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN LATE
CC       MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS MAINTAINED
CC       THROUGH SUBSEQUENT STAGES.
CC   -!- MISCELLANEOUS: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.
CC   -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC       TO AVIDIN/STREPTAVIDIN.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; L07045; AAA30045.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000088; Avidin.
DR   InterPro; IPR000859; CUB_domain.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   Pfam; PF00008; EGF; 8.
DR   Pfam; PF00431; CUB; 1.
DR   Pfam; PF01382; Avidin; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00042; CUB; 1.
DR   SMART; SM00179; EGF_CA; 7.
DR   SMART; SM00001; EGF_like; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 8.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS00577; AVIDIN; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 6.
KW   Biotin; EGF-like domain; Repeat; Signal; Glycoprotein;
KW   Calcium-binding.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    570       FIBROPELLIN C.
FT   DOMAIN       18     55       EGF-LIKE 1.
FT   DOMAIN       62    175       CUB.
FT   DOMAIN      176    212       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      214    250       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      252    288       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      290    326       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      328    364       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      366    402       EGF-LIKE 7.
FT   DOMAIN      404    440       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      442    570       AVIDIN-LIKE.
FT   DISULFID     23     34       BY SIMILARITY.
FT   DISULFID     28     43       BY SIMILARITY.
FT   DISULFID     45     54       BY SIMILARITY.
FT   DISULFID    180    191       BY SIMILARITY.
FT   DISULFID    185    200       BY SIMILARITY.
FT   DISULFID    202    211       BY SIMILARITY.
FT   DISULFID    218    229       BY SIMILARITY.
FT   DISULFID    223    238       BY SIMILARITY.
FT   DISULFID    240    249       BY SIMILARITY.
FT   DISULFID    256    267       BY SIMILARITY.
FT   DISULFID    261    276       BY SIMILARITY.
FT   DISULFID    278    287       BY SIMILARITY.
FT   DISULFID    294    305       BY SIMILARITY.
FT   DISULFID    299    314       BY SIMILARITY.
FT   DISULFID    316    325       BY SIMILARITY.
FT   DISULFID    332    343       BY SIMILARITY.
FT   DISULFID    337    352       BY SIMILARITY.
FT   DISULFID    354    363       BY SIMILARITY.
FT   DISULFID    370    381       BY SIMILARITY.
FT   DISULFID    375    390       BY SIMILARITY.
FT   DISULFID    392    401       BY SIMILARITY.
FT   DISULFID    408    419       BY SIMILARITY.
FT   DISULFID    413    428       BY SIMILARITY.
FT   DISULFID    430    439       BY SIMILARITY.
FT   CARBOHYD     30     30       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    357    357       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   570 AA;  61116 MW;  BE665E3E1C05E6EE CRC64;

Alignment Scores:
Pred. No.:              24.2           Length:        570
Score:                  8.00           Matches:       8
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            1.13%          Indels:        0
```

```
DB:                       1              Gaps:          0

US-10-008-355-1 (1-2139) x FBP3_STRPU (1-570)

Qy    1381 TGTCGGCAGGGATACGCCGGCGTA 1358
           ||||||||||||||||||||||||
Db     430 CysArgGlnGlyTyrAlaGlyVal 437

Search completed: May 16, 2003, 13:07:56
Job time : 48.5 secs
```

```
                    GenCore version 5.1.4_p5_4578
           Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         May 23, 2003, 01:38:11 ; Search time 3816 Seconds
                                         (without alignments)
                                         16313.129 Million cell updates/sec

Title:          US-10-008-355-1
Perfect score:  2139
Sequence:       1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters:      4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*

      Pred. No. is the number of results predicted by chance to have a
          score greater than or equal to the score of the result being printed,
          and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	129.4	6.0	10689	1	AE004008	AE004008 Xylella f
	2	61.6	2.9	10811	1	AE011732	AE011732 Xanthomon
c	3	43.2	2.0	125020	9	AF429315	AF429315 Homo sapi
c	4	42.6	2.0	820	14	AF403408	AF403408 Golden sh
	5	42.2	2.0	1461	3	AY051790	AY051790 Drosophil
c	6	42.2	2.0	76748	2	AC014256	AC014256 Drosophil
	7	42.2	2.0	148847	3	AC008234	AC008234 Drosophil
	8	42.2	2.0	162593	3	AC008356	AC008356 Drosophil
	9	42.2	2.0	224400	3	AE003682	AE003682 Drosophil
	10	41.6	1.9	720	8	CNS01DBA	AL116606 Botrytis
	11	41	1.9	1700	1	RHMNTRBC	M15810 R.meliloti
	12	41	1.9	329100	1	SME591787	AL591787 Sinorhizo
	13	40.4	1.9	125020	9	AF429315	AF429315 Homo sapi
	14	40	1.9	4314	5	AF051784	AF051784 Xenopus l
	15	40	1.9	202306	2	AC121850	AC121850 Mus muscu
	16	39.4	1.8	3060	8	EMEFLBAA	L24395 Emericella
	17	38.8	1.8	348450	1	MLEPRTN4	AL583920 Mycobacte
c	18	38.6	1.8	13155	1	PDSOX	X79242 Paracoccus
	19	38.6	1.8	155962	9	AC099055	AC099055 Homo sapi
c	20	38.4	1.8	100986	8	AC105363	AC105363 Oryza sat
	21	38.4	1.8	126637	2	AC118980	AC118980 Oryza sat
	22	38.2	1.8	190050	1	AL646080	AL646080 Ralstonia
	23	37.8	1.8	31896	2	AC017879	AC017879 Drosophil
c	24	37.8	1.8	172904	3	AC007414	AC007414 Drosophil
c	25	37.8	1.8	275390	3	AE003831	AE003831 Drosophil
c	26	37.4	1.7	38408	2	AC019786	AC019786 Drosophil
	27	37.4	1.7	146153	3	AC007452	AC007452 Drosophil
c	28	37.4	1.7	183213	10	AL669952	AL669952 Mouse DNA
c	29	37.4	1.7	195533	2	AC101759	AC101759 Mus muscu
c	30	37.2	1.7	133691	9	AC074347	AC074347 Homo sapi
c	31	37.2	1.7	170398	2	AC013540	AC013540 Homo sapi
	32	37.2	1.7	329709	1	AP002997	AP002997 Mesorhizo
	33	37	1.7	82746	1	AF453501	AF453501 Actinosyn
c	34	36.8	1.7	810	6	AX110931	AX110931 Sequence
	35	36.8	1.7	2016	8	PAN296102	AJ296102 Podospora
c	36	36.8	1.7	10925	1	AE004750	AE004750 Pseudomon
	37	36.8	1.7	137037	2	AC104321	AC104321 Oryza sat
c	38	36.8	1.7	307150	1	CNSPAX01	AJ248283 Pyrococcu
c	39	36.8	1.7	349980	6	AX041106	AX041106 Sequence
	40	36.6	1.7	1129	8	AF178952	AF178952 Zea mays
	41	36.6	1.7	1496	8	CRENOLM	X66412 C.reinhardt
c	42	36.6	1.7	6709	14	MOCDRNAP	L38927 Molluscum c
c	43	36.6	1.7	95727	2	AC091231	AC091231 Homo sapi
c	44	36.6	1.7	170630	2	AC016905	AC016905 Homo sapi
c	45	36.6	1.7	185097	2	AC023148	AC023148 Homo sapi

ALIGNMENTS

```
RESULT 1
AE004008/c
LOCUS       AE004008                10689 bp    DNA     linear   BCT 15-JUN-2001
DEFINITION  Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.
ACCESSION   AE004008 AE003849
VERSION     AE004008.1  GI:9106961
KEYWORDS    .
SOURCE      Xylella fastidiosa 9a5c.
  ORGANISM  Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 10689)
  AUTHORS   Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
```

```
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
   PUBMED   10910347
REFERENCE   2  (bases 1 to 10689)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
            Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
            13083-970, Brazil
FEATURES             Location/Qualifiers
     source          1. .10689
                     /organism="Xylella fastidiosa 9a5c"
                     /db_xref="taxon:160492"
                     /clone="9a5c"
     gene            complement(119. .367)
                     /gene="XF1871"
     CDS             complement(119. .367)
                     /gene="XF1871"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84677.1"
                     /db_xref="GI:9106962"
                     /translation="MTMTLKEDKTEIIDEQRIQIAKLLSETAKIQAEIHEVNAHTQKL
                     MKETLKLTTESKWYPVVVGSSLIAAGATAATLFIKLFH"
     gene            complement(430. .618)
                     /gene="XF1872"
     CDS             complement(430. .618)
                     /gene="XF1872"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84678.1"
                     /db_xref="GI:9106963"
                     /translation="MMEEARATAPQGGLAASDPQRITDAPKKTGLRKRKYPTTGSASI
                     IVLRHVLTQRQQKRKASP"
     gene            941. .1288
                     /gene="XF1873"
     CDS             941. .1288
                     /gene="XF1873"
                     /note="similar to GI|479340 (percent identity: 36 %/query
                     alignment coverage:  87.8 %/subject alignment coverage:
                     90.2 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF84679.1"
                     /db_xref="GI:9106964"
                     /translation="MSKNKIPNQILTGNWAGFQFINGKLVTPERRTIEEWQLRWLSLT
                     CTLAREWQKMMEEARATAPQGGLAASDPQRITDALQKTVSHKRRHAANETASIIQLRQ
                     VLIQRQQKRKASP"
     gene            1351. .1662
                     /gene="XF1874"
     CDS             1351. .1662
                     /gene="XF1874"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84680.1"
                     /db_xref="GI:9106965"
                     /translation="MTMTLKEDKTEIIDEQKIQIAKLISETAKIQAEIHEVTTHTQEM
                     NAHINEMNARTQEMNARTQKLIEETQKVSKESQWYPVVVASGLLAAGATAATLFIKLF
                     H"
     gene            complement(1839. .2918)
                     /gene="XF1875"
     CDS             complement(1839. .2918)
                     /gene="XF1875"
                     /note="similar to SP|P03626 (percent identity: 36 %/query
                     alignment coverage:  88.0 %/subject alignment coverage:
                     105.0 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="phage-related protein"
                     /protein_id="AAF84681.1"
                     /db_xref="GI:9106966"
                     /translation="MLNLVTGAPGNGKTLYAVDWLIRQIEIDKSLVKSGAVPRSFYTD
                     IEGFDVEAVRRLTGYVVQSAPEDWRTTPQGSVIVYDEAHRMFPAGRPGRSDDPRVCDL
                     DTHRHGGYDLMFVTQWPTKIHHELRRLVGEHVHLNRAMGLQTAGLYRWSRAQDDPYDI
                     HQREKAEEEVWKFPKDRYALYASSTLHTVSHKFRIPKKVWSALSVCVTCSVIGLVFWH
                     YYSPAHLSEAASSVAGASGQASLRAAPASLSSSRSLVSGMRTYAVLETESAPTLSGCV
                     SSDRSCRCFNTDGYQIDMSVVECRRLLASPLPFNVYHAYVTSSSSAAAVSSSPSSSPS
                     SSSVSLSSSVSSAPILGSSSVPSSH"
     gene            complement(2918. .3202)
                     /gene="XF1876"
     CDS             complement(2918. .3202)
                     /gene="XF1876"
                     /note="similar to SP|P03625 (percent identity: 36 %/query
                     alignment coverage:  85.1 %/subject alignment coverage:
                     86.0 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="phage-related protein"
                     /protein_id="AAF84682.1"
                     /db_xref="GI:9106967"
                     /translation="MFSWLAMLLRNILGQTVARGLVGAGLALVTTVPLIPLVTSALNL
                     IVSKMSGISADVLNIALLMGFGEALSIIGSAMLTRLALQSLHVGIVKATT"
     gene            complement(3195. .4412)
                     /gene="XF1877"
     CDS             complement(3195. .4412)
                     /gene="XF1877"
                     /note="hypothetical protein; identified by sequence
```

```
                 similarity; putative; ORF located using Glimmer/RBSfinder"
                 /codon_start=1
                 /transl_table=11
                 /product="hypothetical protein"
                 /protein_id="AAF84683.1"
                 /db_xref="GI:9106968"
                 /translation="MCSRSFFLYFISALGLTLVISFPFCLFAQNVPPTVEVVPTSVRY
                 SSVISDGVSTSATFEARSTALVNGVRYYTVPVDISASTLGSLAKVAVRRGMAFYNVYS
                 MLSGLINGAGWVIDELTHEVMSGPALKEIPVGTVAWWWQRPGDGHVFYSVTPQGLIAP
                 INAYQSSLSPPQPLVTSSGPSSVSSERWLYHLEGGGEYIGHLSKTDQSVPDYSSGLPP
                 SVILDTDLGQLVRSDPSTVNAVLTDSQTGAVLLTPEIVSALNKLRRSLEDELKASHAP
                 DQQPSSGGASSPPSSSGTAWPSFCSWASVVCDFIDWVKSDEFLKKPLVPPDVPYVDKL
                 PQAKTWSSGLGEGACPSPTTIPIEFSGYKTSVEISYQPFCDFAALMRPVVIVIATILA
                 AYIAGGFRGVKNV"
     gene        complement(4481..4702)
                 /gene="XF1878"
     CDS         complement(4481..4702)
                 /gene="XF1878"
                 /note="hypothetical protein; identified by sequence
                 similarity; putative; ORF located using Glimmer/RBSfinder"
                 /codon_start=1
                 /transl_table=11
                 /product="hypothetical protein"
                 /protein_id="AAF84684.1"
                 /db_xref="GI:9106969"
                 /translation="MKSYIDRGIAFFSLLSTSSLVFAADAAGASFDAGAAVTALGAIS
                 GAVALIGAAKLAPAAISVGWKWLKGAIFG"
     gene        complement(4724..4924)
                 /gene="XF1879"
     CDS         complement(4724..4924)
                 /gene="XF1879"
                 /note="hypothetical protein; identified by sequence
                 similarity; putative; ORF located using Glimmer/RBSfinder"
                 /codon_start=1
                 /transl_table=11
                 /product="hypothetical protein"
                 /protein_id="AAF84685.1"
                 /db_xref="GI:9106970"
                 /translation="MAQQLMSLYCTQYDVEARTCSQQAWMVPPSLLPPISYEDVRILL
                 PHIVMCFLVAWGFHFLFTVVRD"
     gene        complement(4931..5236)

  Query Match             6.0%;  Score 129.4;  DB 1;  Length 10689;
  Best Local Similarity   44.9%;  Pred. No. 3.1e-23;
  Matches  888;  Conservative    0;  Mismatches 1051;  Indels   38;  Gaps    9;

Qy      188 CCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCC 247
Db    10327 CCATGGGGCAGTGGTCTCTCTCGGCAATTGCACCGCGAGTCTGGTTTCACCAGAAGGAC 10268

Qy      248 TGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATC 307
Db    10267 TGGTGATCACTAACCATCATTGCGCTTATGGTGCGATTCAGTTGAACTCTACGCCTAAAA 10208

Qy      308 ACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTT----CCGATT 363
Db    10207 AGAACTTAATCAAAGAAGGGTTCAATGCGTTGACTCAGGCTGATGAAGTCAGTGCGGGTC 10148

Qy      364 CCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGT--AGAAG 421
Db    10147 CGAATGCGCGTATTTACGTGCTTGAGCAGATTACCGATGTCACCGCTCAAGCTAAAGCTG 10088

Qy      422 GACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCC 481
Db    10087 CTCTGGCTGCTGCCGGTAATGACCCCTTCAAGCGTACGACGGCATTGGAGACATTCAGCA 10028

Qy      482 AAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCT 541
Db    10027 AACAAGAAATCGCCAAATGCGAGGAAGAGCAGGGTTATCGCTGCCAATTCTTCAGTTTCG 9968

Qy      542 ATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTAT 601
Db     9967 CTGGTGGAAATACCTACAGAGTGTTCAAGAACCTGGAGATCAAGGACGTGCGTTTGGTTT 9908

Qy      602 TTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGC 661
Db     9907 ACGCACCGCAGGGGAGTGTGGGTAAGTTCGGGGGCGACGTCGATAACTGGATGTGGCCGC 9848

Qy      662 GTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCG 721
Db     9847 GTCATACCGGTGACTTTTCTTTCTACCGTGCCTACGTCGGTAAAGACGGTAAACCAGCGA 9788

Qy      722 AATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAG 781
Db     9787 GCTTCTCTAAAGAAAACATACCCTACCGTCCCAAACACTGGCTGAAATTTAGCGATCAAC 9728

Qy      782 GCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACC 841
Db     9727 CATTAGGTGATGGTGACTTCGTGATGGTCGCTGGTTACCCAGGACGTACCAATCGTTATG 9668

Qy      842 TCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTC 901
Db     9667 CTTTGGTGGCCGAATTTGAGAATACTGCTCACTGGACTTATCCGGTGATTGGCCAGCATT 9608

Qy      902 GCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCA 961
Db     9607 TCAAGAATCTCATTGCTTTGATTGAGGCGGCTAGCAAACAGAATCCAGACATTCAAGTGA 9548

Qy      962 AATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACC 1021
Db     9547 AATATGCGAGCACTTTAGCTGGATTGAATAACACATCGAAAACTTTGATGGTCAATTGG 9488

Qy     1022 GCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAG 1081
Db     9487 ATGGCTTCCGACGCATTAATGCTATAGGTCAGAAACAGAGTGAAGAAACAGCAGTATTAG 9428

Qy     1082 ACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAA 1141
Db     9427 CTTGGTTGAAACAGCAGGGCATACGTGGACACGAAGCATTGGCTGCGCATCAAACATTGG 9368

Qy     1142 AGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCT 1201
Db     9367 TTGATCTGACAGAACAATACAAAGCAAATCAGGATCGTGATTTTGTATTGGGTCAGTTCA 9308

Qy     1202 TCG---GTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATC 1258
Db     9307 ACGGAAGTGGTGTCATTGGTGTAGCGGTGAACCTATATCGCTTGGCAATCGAGCGCACAA 9248

Qy     1259 CTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCT 1318
Db     9247 AATCTGATGCCCAACGTGAAGCGGGATATCAGGAGCGCGATCTGCCAACCATTGAAGGCA 9188

Qy     1319 CGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCG 1378
Db     9187 ACTTAAAGCAGATGGAGCGACGTTACTTACCGGAAATGGATCGCCAG---ATGCAGCAAT 9131

Qy     1379 ACAAGCTCCCCGATATATTCAAGAATGTAATCGACAA----------GAAATTCAAAGGC 1428
Db     9130 ACTGGCTCACCGAATACAACAAGCTTCCTGTCAAACAGCGTGTGGCTGCAATTGATGTTT 9071

Qy     1429 GACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAG 1488
Db     9070 GGTTGGGAGATGGCATCCCAGCAACTCTCAAGCGCTTGGGTGACACCAAATTAAG-CAGT 9012

Qy     1489 TTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGAT 1548
Db     9011 TCTGAGGAAAGGCTGAAGTGGTTTAATGCCGACCGTGCAGCCTTCGAAAGCAGTCAAGAT 8952

Qy     1549 CCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCG 1608
Db     8951 CCAGCGATCCGTTATGCTGTGGCCATCATGCCAGCATTACTGGAGATTGAACGTCAGAAT 8892

Qy     1609 ATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAG 1668
Db     8891 AAAATCCGTACTGGAGAACTACTTAAAGCGCGTCCGATCTACTTGCAAGCCTTAGCTGAC 8832
```

```
Qy    1669 ATGTACCCCGGACGTG------CTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGC 1722
              ||    | | ||         | |        || |||||||   |  | ||||| | |
Db    8831 TACAACAAGAGTCATGGCAAATTTGTTTACCCAGACGCCAACTCATCATTACGTATTACC 8772

Qy    1723 TACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACA 1782
           | ||||  | |  || || |||   |||   | || || |  |   |||||       |||||||
Db    8771 TTCGGCCACGTGAAAGGCTATTCACCCAAAGATGGCGTCGAGTACACACCGTTTACGACA 8712

Qy    1783 GGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAAT 1842
               |||| ||  ||| ||   || || |     |||     || |    || | ||
Db    8711 CTGCAGGGTGTCATGGCGA---AGAATACAGGCGTCGAGCCATTCGATTCACCGAAGTCA 8655

Qy    1843 ATCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGA------GAACGGTCAG 1896
             |  |  |      |      | || | ||| |      |  |||||        |  ||
Db    8654 CTGATTAATGCGATAAAGGCAAAGAGCTACGCCAATTTAGCCGATCAGCGTATTGGCACT 8595

Qy    1897 CTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTA 1956
             ||| |  |    ||| |||||  ||    ||||| || ||||| ||||||||     || ||
Db    8594 GTCCCTGTGAATTTTCTATCTGACTTGGACATTACTGGCGGCAACTCCGGATCGCCAGTC 8535

Qy    1957 TTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGT 2016
           || |||       ||||      || | ||| | || || || || |||||  || | |||
Db    8534 TTAGATGCCCACGGAAAATTGGTGGGTTTAGCATTTGACGGTAATTGGGAGTCTGTCAGT 8475

Qy    2017 GGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTT 2076
             || |      | ||| || || |     ||     || |||||      || || |   || || ||
Db    8474 AGTAATTGGGTGTTTGATCCTGTGATGACTCGAACAATTGCAGTTGATAGTCGTTATGTG 8415

Qy    2077 CTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTG 2133
             |  |  || ||       ||    |   |      || | |||   |   | ||   |||| |||
Db    8414 CAGTGGATCATGACTGAAGTTGCACCAGCACCACATCTACTAAAGGAATTGAATTTG 8358


RESULT 2
AE011732
LOCUS       AE011732                10811 bp    DNA     linear   BCT 29-MAY-2002
DEFINITION  Xanthomonas axonopodis pv. citri str. 306,  section 110 of 469 of
            the complete genome.
ACCESSION   AE011732 AE008923
VERSION     AE011732.1  GI:21107161
KEYWORDS    .
SOURCE      Xanthomonas axonopodis pv. citri str. 306.
  ORGANISM  Xanthomonas axonopodis pv. citri str. 306
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xanthomonas.
REFERENCE   1  (bases 1 to 10811)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Comparison of the genomes of two Xanthomonas pathogens with
            differing host specificities
  JOURNAL   Nature 417 (6887), 459-463 (2002)
   MEDLINE  22022145
    PUBMED  12024217
REFERENCE   2  (bases 1 to 10811)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2001) Departmento de Bioquimica, Universidade de
            Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
            Brazil
FEATURES             Location/Qualifiers
     source          1..10811
                     /organism="Xanthomonas axonopodis pv. citri str. 306"
                     /strain="306"
                     /db_xref="taxon:190486"
                     /note="pathovar: citri"
     gene            complement(293..1315)
                     /gene="tdh"
                     /note="XAC1022"
     CDS             complement(293..1315)
                     /gene="tdh"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="threonine 3-dehydrogenase"
                     /protein_id="AAM35905.1"
                     /db_xref="GI:21107162"
                     /translation="MKALVKREANKGIWLEQVPVPTPGPNEVLIKLEKTAICGTDLHI
                     YLWDEWSQRTIEPGLTIGHEFVGRVAELGSAVTGYQIGQRVSAEGHIVCGHCRNCRGG
                     RPHLCPNTVGIGVNVNGAFAEYMVMPASNLWPIPDQIPSELAAFFDPYGNAAHCALEF
                     DVIGEDVLITGAGPIGIIAAGICKHIGARNVVVTDVNDFRLKLAADMGATRVVNVSKT
                     SLKDVMADLHMEGFDVGLEMSGNSRAFNDMLDCMYHGGKIAMLGIMPRGAGCDWDKII
                     FKGLTVQGIYGRKMYETWYKMTQLVLSGFPLHKVLTHQLPIDDFQKGFDLMEEGKAGK
                     VVLSWN"
     gene            complement(1840..4260)
                     /gene="fecA"
                     /note="XAC1023"
     CDS             complement(1840..4260)
                     /gene="fecA"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="TonB-dependent receptor"
                     /protein_id="AAM35906.1"
                     /db_xref="GI:21107163"
                     /translation="MFHEPLTPCVAKRRLCTVRPPSAVPEQDSPMTSARRFHVTPLAL
                     ALATLLPVISQAATPADPAAVAPAEAALTADAAANATQLDAINVVSQGSTRQVQRISQ
                     QDIEQLTPGSSAFKAVEKLPGVQFQSADPFGTYEWSTQVTLHGFDQSRLGYTLDGIPL
                     GNMSYGVTNGLHITRAIISENLGSVEIAQGAGALGTASNTNLGGTMQFYSADPQTTPG
                     ARLVQTVGSDATRRTFVRADTGDRDGLSAYLSYANASTDKWKGYGDQTSEQANLKTVY
                     QWGDGNRLSLFVDTSRRKEYDYMDLSLTSQRALGWNYDYLQPDWATAVQMARAYQNTG
                     ATSGVANGYPQSLAGLPSDYSWLDASYYAGGGLRRDNLAGLSGTFVFGGATLDASGYY
                     HGNRGEGQWVTPYVRTSAQIPVSMRTTDYGLDRFGGTSALKWSWGNHDLEVGAWAENA
                     RTTQGRNYFALGNEGYTSLYNVYEAQTPFRRDFLQRYTTQTRMLYVQDTVRLLDDRLT
                     LNYGAKALSTTTRAQSLVPTTALAQGRIRAEDNFLPQLGINYKLDERQDLYASYSKNI
                     AAFGFTPFATSQAAFDRSRSTLEPEQSQTVQVGYRVQDAQFQLSADAYFTKFSNRLLT
                     TSPCTAVQTCASILSNVGAVHSAGADLALMWRPTEGLSWLNSLSWNRSRYQDDYLNNG
                     VVATSGKDVVGIPALMFSSSASYQIGNLRLDLDGKYVDKRYITFLNDSQVPSYWLFNA
                     GARYDFGRVGGVADVALALNISNLTDKRYFASTGTNGYVASDPDGYNQTMVVGAPRQY
                     FMSLDVKF"
     gene            4369..6450
                     /gene="XAC1024"
     CDS             4369..6450
                     /gene="XAC1024"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
```

```
                    /codon_start=1
                    /transl_table=11
                    /product="non-hemolytic phospholipase C"
                    /protein_id="AAM35907.1"
                    /db_xref="GI:21107164"
                    /translation="MTSSSSRRDFLKRVAALTAAGALPSSIGRALALPANARTGTLRD
                    LEHVVILMQENRSFDHYFGALRGVRGFGDPRALQLRDGHAVWSQPTPDGRRVLPFAFD
                    SQNTCAPLIKSLDHSWKAGHGQDPARWAEYDAWVPYKGELTMGYFQRHDIPYYHALAD
                    AFTICDGYYCSLHGPTNPNRMYLFTGTSGPSVGNVGAQAVTNADDGNWTADMARDKPG
                    YAALQWTTYAQRLQAAGVDWRVYQEYDNFGCNSLAYFSHYRDLRTDDERYLRARACVP
                    GSTAGNAATTQAEHLVAAIAADVRANRLPQVSWVIPPTAYCEHPEAPPAYGESLVARL
                    IDALTANPEVWAKTALIINYDENDGFFDHVPAPLPALDARMGRSNVDTHGEVYDGVPI
                    GLGIRVPMLVISPWTRGGWVNSQVFDHTSVLRLLERRFGVAEPNISPWRRAVSGDLTS
                    VFDFRKPDDSALSALPSVDDYRARTAAVRDKPLPSAPAAATMPRQEPGQRPARALPYA
                    LQVHARVQEGAAVQLQFVNSGAAAAAFNVYNSAAGGGPWYYTVLPGTQLDDAPIGATH
                    EGAYALRVHGPNGFLREFAGQVPRASAPSAAPWVEARQEGDALVLEIGNAGQRACTVQ
                    LRALDYADPSARTLSLAAGQRETIRLALAASDHWYDLVVEQPGSAFRRRLAGHLETGR
                    PSRSDPAFGRV"
     gene           6617..6766
                    /gene="XAC1025"
     CDS            6617..6766
                    /gene="XAC1025"
                    /note="identified by sequence similarity; putative; ORF
                    located using Blastx/Glimmer/Genemark"
                    /codon_start=1
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAM35908.1"
                    /db_xref="GI:21107165"
                    /translation="MRPNLFAVSVLATLAVAASAQAGEGMWVPQQLPEIAGPPQQAGL
                    RLSPR"
     gene           6791..7054
                    /gene="XAC1026"
     CDS            6791..7054
                    /gene="XAC1026"
                    /note="identified by sequence similarity; putative; ORF
                    located using Blastx/Glimmer/Genemark"
                    /codon_start=1
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAM35909.1"
                    /db_xref="GI:21107166"
                    /translation="MGAVVALGGCTASFVSPQGLVVTNHHCAYGAIQLNSTAQKNLIK
                    DGFNAVRPADELSAGPSARIYVLDAITDVTAPAKAAMATPVRR"
     gene           7051..7236
                    /gene="XAC1027"
     CDS            7051..7236
                    /gene="XAC1027"
                    /note="identified by sequence similarity; putative; ORF
                    located using Blastx/Glimmer/Genemark"
                    /codon_start=1
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAM35910.1"
                    /db_xref="GI:21107167"
                    /translation="MMDAQGKLVGLAFDGNWESVSSNWIFDPAMTRMIAVDGRYLRWI
                    MTEVAPAPQLLKELGVR"
     gene           7329..7973
                    /gene="pgmA"
                    /note="XAC1028"
     CDS            7329..7973
                    /gene="pgmA"
                    /note="identified by sequence similarity; putative; ORF
                    located using Blastx/Glimmer/Genemark"
                    /codon_start=1
                    /transl_table=11
                    /product="phosphoglycerate mutase"
                    /protein_id="AAM35911.1"
                    /db_xref="GI:21107168"
                    /translation="MRILLARHGETPWNAEGRYQGQIDIPLSPVGEGQARALGERLHS
                    LQIDRAVASPLSRAQATAKAALGASRQALLQTDADLQEIAHGEWEGLLASEINDKDPA
                    RLRAWREEPDTVLMPGGESLRQVLDRSWRGLMRAADGLGAHDTLLVVAHDAVNRVILC
                    KILGLPLSRLWSFRQAPTTLNLLEGDDVEHLEVVRLNDCAHHTPFFGEAKHRAL"
     gene           8242..9540
                    /gene="folC"
                    /note="XAC1029"
     CDS            8242..9540
                    /gene="folC"
                    /note="identified by sequence similarity; putative; ORF
                    located using Blastx/Glimmer/Genemark; dihydrofolate
                    synthase"
                    /codon_start=1
                    /transl_table=11
                    /product="folylpolyglutamate synthase"
                    /protein_id="AAM35912.1"
                    /db_xref="GI:21107169"
                    /translation="MNATDSLSAWLAYIEQQHPSAIAMGLERVREVAVRLQIAAPAKH
                    VIVVGGTNGKGSTVAFIEAIGRAAGWKVGAYTSPHLLRYNERVRIDGQEASDAQLVDA
                    FAAVEAARGDTALTYFEFGTLAALWLFQQSALELAVLEIGLGGRLDAVNIVDSDVAVI

  Query Match             2.9%;  Score 61.6;  DB 1;  Length 10811;
  Best Local Similarity   56.4%;  Pred. No. 5e-05;
  Matches  115;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy   1936 GGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGAT 1995
          || |||||||||   || ||  | ||   | | |||   ||| |||| || || |||||
Db   7033 GGCAACTCCGGTTCGCCGGTGATGGACGCGCAGGGCAAGCTGGTCGGCCTGGCGTTCGAC 7092

Qy   1996 GGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATC 2055
          || || |||||  |  ||||  |  ||      ||||| || |   ||   ||||  |||
Db   7093 GGAAATTGGGAGTCGGTGAGCAGCAACTGGATCTTCGACCCGGCAATGACCCGCATGATC 7152

Qy   2056 AGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGTCAGTGCCCCCGTCTC 2115
            |||||||  ||||||  | |  |  || ||   | |    | || |   || |  ||
Db   7153 GCCGTGGACGGCCGCTATCTGCGTTGGATCATGACCGAGGTCGCTCCGGCGCCACAGCTG 7212

Qy   2116 ATCCAAGAGCTGAAGTTGATCTAA 2139
           |  | || ||     ||  || |
Db   7213 TTGAAGGAACTCGGCGTGCGCTGA 7236


RESULT 3
AF429315/c
LOCUS       AF429315           125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
   MEDLINE  21583737
    PUBMED  11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1..125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581..35746)
                     /rpt_type=tandem
```

```
                     /rpt_unit=ctg
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT     29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN

  Query Match             2.0%;  Score 43.2;  DB 9;  Length 125020;
  Best Local Similarity   9.8%;  Pred. No. 5.9;
  Matches   46;  Conservative  226;  Mismatches  194;  Indels    4;  Gaps    1;

Qy        250 ATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCAC 309
Db      51261 AKCWWWYSYRKYCWMSMMMASWAAMAMWMMCTAAWYYKKWTRMRRWSKYSSSYAYAWYAT 51202

Qy        310 GACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGT 369
Db      51201 MCWYCYSSMYAARRMGTYSGGKKRKWASTSCMCCCRRGCCMKKKGMRWCSTTYTGMSGGK 51142

Qy        370 CTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTC 429
Db      51141 SRWWTSSMACMYWYKGSRRKTKSMATSGCMRMGAMRSKGGMRKYWSCRYKGMRWGRWSGY 51082

Qy        430 AAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTG 489
Db      51081 KGCSGYMSMGMYGSGRTSKYCSYTGGYCSKCSAKMCKSKSTSKSCCTKSKYS----CSMG 51026

Qy        490 GCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAAC 549
Db      51025 RYSSYCTCWGGRMCKGMKSSCCASRSGCKSYKMAGSWRCRGRRGMSCWYWSSMKSYSMAS 50966

Qy        550 AACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCT 609
Db      50965 YKSSARGRGMKCYKTSCMKYYYSMHBKSSHDBSGYKSYRKSYBYWTCMMKDSMCWSVDSH 50906

Qy        610 CCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACG 669
Db      50905 MMRDMMYDKMRMKVSSGTRGYHBHSAWMCKSMSRDSDSGYSGMYHMKSYWKSSASKHMYY 50846

Qy        670 GGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGC 719
Db      50845 SGMKSSVYBMDCMSMYTSBSKSVRMBRRWSGSYBRYRKCAVMRRYBHRSS 50796


RESULT 4
AF403408/c
LOCUS       AF403408               820 bp ds-RNA    linear   VRL 07-AUG-2002
DEFINITION  Golden shiner reovirus segment 11 non-structural protein NS3 gene,
            complete cds.
ACCESSION   AF403408
VERSION     AF403408.1  GI:22128470
KEYWORDS    .
SOURCE      Golden shiner reovirus.
  ORGANISM  Golden shiner reovirus
            Viruses; dsRNA viruses; Reoviridae; Aquareovirus.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   Attoui,H., Fang,Q., Jaafar,F.M., Cantaloube,J.F., Biagini,P., De
            Micco,P. and De Lamballerie,X.
  TITLE     Common evolutionary origin of aquareoviruses and orthoreoviruses
            revealed by genome characterization of Golden shiner reovirus,
            Grass carp reovirus, Striped bass reovirus and golden ide reovirus
            (genus Aquareovirus, family Reoviridae)
  JOURNAL   J. Gen. Virol. 83 (Pt 8), 1941-1951 (2002)
   PUBMED   12124458
REFERENCE   2  (bases 1 to 820)
  AUTHORS   Attoui,H., de Micco,P. and de Lamballerie,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2001) Virologie, Faculte de Medecine de
            Marseille, 27 Boulevard Jean Moulin, Marseille 13005, France
FEATURES             Location/Qualifiers
     source          1. .820
                     /organism="Golden shiner reovirus"
                     /db_xref="taxon:185783"
                     /note="Golden shiner reovirus
                     segment: 11"
     CDS             43. .777
                     /codon_start=1
                     /product="non-structural protein NS3"
                     /protein_id="AAM92755.1"
                     /db_xref="GI:22128471"
                     /translation="MATAGITAAIDDLVISRIDEAITNNTLVQRLRKKVSSLWKDYDA
                     HVSDPTAHELLTMPAALADLPPPPPPTDDEINRSFYATALPSVVPTSDTLWGTASYEI
                     IGRAVRIASSDKAPRLILHVTDETIRLEVPTMSFTLASPLSTGSLLRVTLPRGAIHSS
                     LTGKFDERCWVRLLSTLTKWSTPSTMVTSDWTLDTTTGLTKPPVTGFNARATVKDSAI
                     VMWFPAPDIATSARRAIVSPMSCVFE"
BASE COUNT      161 a    288 c    196 g    175 t
ORIGIN

  Query Match             2.0%;  Score 42.6;  DB 14;  Length 820;
  Best Local Similarity   54.0%;  Pred. No. 4.6;
  Matches   87;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy       1858 CGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCG 1917
Db        744 CGCCCGACGAGCGGAGGTTGCGATGTCGGGAGCGGGGAACCACATAACGATGGCGGAGTC 685

Qy       1918 AACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTG 1977
Db        684 TTTAACGGTCGCACGGGCGTTGAATCCAGTGACGGGGGGTTTGGTCAGACCGGTGGTGGT 625

Qy       1978 ATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGG 2018
Db        624 GTCCAACGTCCAGTCACTGGTGACCATGGTGCTGGGGGTGG 584


RESULT 5
AY051790
LOCUS       AY051790              1461 bp    mRNA    linear   INV 27-AUG-2001
DEFINITION  Drosophila melanogaster LD30634 full length cDNA.
ACCESSION   AY051790
VERSION     AY051790.1  GI:15291890
KEYWORDS    FLI_CDNA.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1461)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
            Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
            Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K.,
            Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2001) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
            This clone was sequenced as part of a high-throughput process to
```

```
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000). The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone.  However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1..1461
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /map="85D8-85D9"
     gene            1..1461
                     /gene="CG8444"
                     /note="alignment with genomic scaffold AE003682"
                     /db_xref="FLYBASE:FBgn0037671"
     CDS             53..1015
                     /gene="CG8444"
                     /note="Longest ORF"
                     /codon_start=1
                     /product="LD30634p"
                     /protein_id="AAK93214.1"
                     /db_xref="GI:15291891"
                     /db_xref="FLYBASE:FBgn0037671"
                     /translation="MLRVFVIFSLFIAAINASGEFTVLNRPKAISFKGNDALESHYVG
                     DVLYASMGNAVSGDTNWNGLTINDPFNLAKGVILVHVQGIGHVTTAGNVKTYELTGSG
                     TDASLNALAAELEAANEPVCDINFEQFDDGVQAWKSCFGDFEAPAAKPTKHLNPSLHT
                     ADKQFLQEVGFINSAADHLAEMAKPSNVLMLRVSVDGVAKAHGEKSVAVEEANKLLSA
                     AISRLLAASQKSSDSVLFVQTTEKDVAASRAKRDTIAASTTNPYNLAVYYGSDYPVIF
                     NIILWFMVVFGLSLLAICYAIAAMDPGRDSIIYRMTSTRIKKDN"
BASE COUNT      374 a    376 c    352 g    359 t
ORIGIN

  Query Match             2.0%;  Score 42.2;  DB 3;  Length 1461;
  Best Local Similarity   49.3%;  Pred. No. 6.4;
  Matches  110;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy       1470 GGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGC 1529
              ||  ||  |||  |  |  |  ||||  |  |||||||  |||  ||    |||||
Db        703 GGCCGCTATCAGCCGCCTGCTTGCTGCCTCCCAGAAGTCCAGCGACTCGGTTCTGTTTGT 762

Qy       1530 CAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
              | || |  | ||||| ||  ||||| |  || | ||||| | |  |    ||| |||
Db        763 CCAGACCACGGAGAAGGACGTGGCAGCATCGCGTGCCAAGCGTGATACCATTGCCGCTTC 822

Qy       1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
              | | |  |  || |    |||||    ||| | ||  ||  |      |    |
Db        823 CACCACCAATCCCTACAACCTGGCCGTCTACTACGGCAGCGACTACCCGGTGATCTTCAA 882

Qy       1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCG 1692
              | |   |  | |     | |   | |||||   |||||| |
Db        883 CATCATCCTGTGGTTCATGGTCGTCTTCGGACTGTCTCTGCTG 925


RESULT 6
AC014256/c
LOCUS       AC014256            76748 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014256
VERSION     AC014256.1  GI:6437079
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 76748)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211789 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..76748
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    21897 a  15665 c  15830 g  23356 t
ORIGIN

  Query Match             2.0%;  Score 42.2;  DB 2;  Length 76748;
  Best Local Similarity   49.3%;  Pred. No. 10;
  Matches  110;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy       1470 GGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGC 1529
              ||  ||  |||  |  |  |  ||||  |  |||||||  |||  ||    |||||
Db      17172 GGCCGCTATCAGCCGCCTGCTTGCTGCCTCCCAGAAGTCCAGCGACTCGGTTCTGTTTGT 17113

Qy       1530 CAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
              | || |  | ||||| ||  ||||| |  || | ||||| | |  |    ||| |||
Db      17112 CCAGACCACGGAGAAGGACGTGGCAGCATCGCGTGCCAAGCGTGATACCATTGCCGCTTC 17053

Qy       1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
              | | |  |  || |    |||||    ||| | ||  ||  |      |    |
Db      17052 CACCACCAATCCCTACAACCTGGCCGTCTACTACGGCAGCGACTACCCGGTGATCTTCAA 16993

Qy       1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCG 1692
              | |   |  | |     | |   | |||||   |||||| |
Db      16992 CATCATCCTGTGGTTCATGGTCGTCTTCGGACTGTCTCTGCTG 16950


RESULT 7
AC008234
LOCUS       AC008234               148847 bp    DNA     linear   INV 17-FEB-2001
DEFINITION  Drosophila melanogaster, chromosome 3R, region 85D-85D, BAC clone
            BACR32M04, complete sequence.
ACCESSION   AC008234
VERSION     AC008234.8  GI:12957667
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 148847)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 3R, region 85D-85D
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148847)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
```

```
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 17, 2001 this sequence version replaced gi:6984365. .
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .148847
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="85D-85D"
                     /clone="BACR32M04 (D969)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    43851 a  31005 c  30227 g  43764 t
ORIGIN

  Query Match             2.0%;  Score 42.2;  DB 3;  Length 148847;
  Best Local Similarity   49.3%;  Pred. No. 11;
  Matches  110;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy       1470 GGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGC 1529
              ||  ||  |||  |  | |  ||||  |  ||||||| ||| ||    |||||
Db       7940 GGCCGCTATCAGCCGCCTGCTTGCTGCCTCCCAGAAGTCCAGCGACTCGGTTCTGTTTGT 7999

Qy       1530 CAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
              | || | |  ||||| ||  ||||| |  || | |||||| | |  |   ||| |||
Db       8000 CCAGACCACGGAGAAGGACGTGGCAGCATCGCGTGCCAAGCGTGATACCATTGCCGCTTC 8059

Qy       1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
              | | |   |  || |     |||||     ||| | || || |       |    |
Db       8060 CACCACCAATCCCTACAACCTGGCCGTCTACTACGGCAGCGACTACCCGGTGATCTTCAA 8119

Qy       1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCG 1692
              | |  |  | |      | |   | |||||  |||||| |
Db       8120 CATCATCCTGTGGTTCATGGTCGTCTTCGGACTGTCTCTGCTG 8162


RESULT 8
AC008356
LOCUS       AC008356              162593 bp    DNA     linear   INV 06-SEP-2001
DEFINITION  Drosophila melanogaster, chromosome 3R, region 85D-85D, BAC clone
            BACR43K14, complete sequence.
ACCESSION   AC008356
VERSION     AC008356.5  GI:15451512
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 162593)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 3R, region 85D-85D
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162593)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Sep 6, 2001 this sequence version replaced gi:12957666.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .162593
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="85D-85D"
                     /clone="BACR43K14 (D968)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    49906 a  32795 c  32702 g  47190 t
ORIGIN

  Query Match             2.0%;  Score 42.2;  DB 3;  Length 162593;
  Best Local Similarity   49.3%;  Pred. No. 11;
  Matches  110;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy       1470 GGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGC 1529
              ||  ||  |||  |  | |  ||||  |  ||||||| ||| ||    |||||
Db     123301 GGCCGCTATCAGCCGCCTGCTTGCTGCCTCCCAGAAGTCCAGCGACTCGGTTCTGTTTGT 123360

Qy       1530 CAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
              | || | |  ||||| ||  ||||| |  || | |||||| | |  |   ||| |||
Db     123361 CCAGACCACGGAGAAGGACGTGGCAGCATCGCGTGCCAAGCGTGATACCATTGCCGCTTC 123420

Qy       1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
              | | |   |  || |     |||||     ||| | || || |       |    |
Db     123421 CACCACCAATCCCTACAACCTGGCCGTCTACTACGGCAGCGACTACCCGGTGATCTTCAA 123480

Qy       1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCG 1692
              | |  |  | |      | |   | |||||  |||||| |
Db     123481 CATCATCCTGTGGTTCATGGTCGTCTTCGGACTGTCTCTGCTG 123523


RESULT 9
AE003682
```

```
LOCUS       AE003682            224400 bp    DNA     linear   INV 05-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386035 section 7
            of 105, complete sequence.
ACCESSION   AE003682 AE002708
VERSION     AE003682.2  GI:10726402
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 224400)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
   PUBMED   10731132
REFERENCE   2  (bases 1 to 224400)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7299142.
FEATURES             Location/Qualifiers
     source          1. .224400
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
     gene            complement(<235. .>2417)
                     /gene="D1"
                     /note="CG9745"
                     /map="85D2-85D3"
                     /db_xref="FLYBASE:FBan0009745"
                     /db_xref="FLYBASE:FBgn0000412"
     mRNA            complement(join(235. .1324,1643. .2221,2355. .2417))
                     /gene="D1"
                     /product="CT27541"
                     /db_xref="FLYBASE:FBan0009745"
                     /db_xref="FLYBASE:FBgn0000412"
     CDS             complement(join(672. .1324,1643. .2057))
                     /gene="D1"
                     /note="D1 gene product"
                     /codon_start=1
                     /protein_id="AAF54341.1"
                     /db_xref="GI:7299143"
                     /db_xref="FLYBASE:FBan0009745"
                     /db_xref="FLYBASE:FBgn0000412"
                     /translation="MEEVAVKKRGRPSKASVGGKSSTAAVAAISPGIKKRGRPAKNKG
                     SSGGGGQRGRPPKASKIQNDEDPEDEGEEDGDGDGSGAELANNSSPSPTKGRGRPKSS
                     GGAGSGSGDSVKTPGSAKKRKAGRPKKHQPSDSENEDDQDEDDDGNSSIEERRPVGRP
                     SAGSVNLNISRTGRGLGRPKKRAVESNGDGEPQVPKKRGRPPQNKSGSGGSTGYVPTG
                     RPRGRPKANAAPVEKHEDNDDDQDDENSGEEEHSSPEKTVVAPKKRGRPSLAAGKVSK
                     EETTKPRSRPAKNIDDDADDADSADQGQHNSKKESNDEDRAVDGTPTKGDGLKWNSDG
                     ENDANDGYVSDNYNDSESVAA"
     gene            complement(<3004. .>7907)
                     /gene="CG9746"
                     /map="85D2-85D4"
                     /db_xref="FLYBASE:FBan0009746"
                     /db_xref="FLYBASE:FBgn0037663"
     mRNA            complement(join(3004. .3235,3300. .3914,3972. .4605,
                     4986. .5199,5255. .6265,6330. .7367,7561. .7747,7809. .>7907))
                     /gene="CG9746"
                     /product="CT27547"
                     /db_xref="FLYBASE:FBan0009746"
                     /db_xref="FLYBASE:FBgn0037663"
     CDS             complement(join(3004. .3235,3300. .3914,3972. .4605,
                     4986. .5199,5255. .6265,6330. .7367,7561. .7747,7809. .7906))
                     /gene="CG9746"
                     /note="CG9746 gene product"
                     /codon_start=1
                     /protein_id="AAF54342.1"
                     /db_xref="GI:7299144"
                     /db_xref="FLYBASE:FBan0009746"
                     /db_xref="FLYBASE:FBgn0037663"
                     /translation="MGNQLVGIAPSQIYAVEHYFSGQFGSEIAFSSNMGSTRFFKVAK
                     AKTDEGQIVVKVFVKHDPTLPLEDHKERLEGIKKTLSLANAVNCLPFQRVELIDKAAY
                     IMREYVKHSLYDRVSTRPFLTVLEKKWITFQILCALNQCHKQKICHGDIKLENILITS
                     WNWILLSDFASFKPTYLPEDNPADYTYFFDTSRRRTCYIAPERFVKTLASDDDGGNGN
                     MSVIHTDSIIRLAPSYAGNTLLPAMDIFSAGCALLELWTEGTAPFELSQLLAYRRGER
                     DLVEKHLAGIENERLRNLLASMIDIHSMNRKSAEDYLDQERGQLFPEYFYSFLQSYLQ
                     MFSSTPIMSPDDKIQCLHKDIGHCIKVLTQAQDLTPEEEEKEEKKDPTLCARLPPEHD
                     GLILIITVVTSCIRGLKQSNTKIAALELLQKLSKYTTSETILDRILPYILHLAQKSPA
                     KVQVQAIETLTACLGMVKDIPRSDANVFPEYILPSITDLASETSAVIVRVAFARNIAA
                     LAKSAVYFLEETQRNAPNDMPTPRYEAELNALHDIVRQAVLSLLTDSQPVVKQTLMES
                     GICDLCAFFGKEKANDVILSHIMTFLNDEDKNLRGAFYDNIAGVAGYVGWQASDILVP
                     LLQQGFTDREEFVIAKAIRAVTILIELGHIKKPAVTEIVQETACYLCHPNLWVRHEIC
                     GMIATTARNLSAIDVQCKIMPAIGAFLKAPLIQVEKPHILLDCVHPPVPRQIFDSVLR
                     FQDIHHFIRALEARSRVRSQTRQGALPQYEEMGQTLRNLFRRLSSEGLTDLIEMQLLA
                     MNPFLISMKHKALQDLDDTASGNGRIVVSRKQVACHEYPLADKASKMPLDNRSSEGPT
                     LASPASDAAITPLGAGGLAGSPASGAVVLGVPTATVSAATSLGEYTMPERNCWLERSS
                     ECRKDLESLTAKIKSRYNVLAISQRCSYDKLVTPYPPGWRMTGTLVAHLHEHSESVIK
                     LASLRPHGSLFASGSIDGTVRLWDCSKLNGNQGVNKSRQVYSANTPIYALAACDSGQS
                     LAVGGKDGSMLMMRIDRNSAKMTLQQALNQNDHKDGPVVDMHAYDQATQSVIVYATLY
                     GGIVAWDTRMQHSAWRLQNELRHGVITTICADPTGSWLATGTSGGKHICWDLRFRLPI
                     AEIKHPADSWIRKVACHPTEPSYLISASQSNNEVSVWNIETGQRQTVLWASPVAALSS
                     ASPSDPATTCGILSGVVDGSPFILTGSSDQRIRYWDITNPKNSSLKVPAANDNLADVA
                     FNYGARLIDGSQVIEEQIISMASTGDSQQLRSSTEENPRSGPDMPTASHHDAITDLLM
                     CKDKGQIYIASASRNGVIKLWK"
     gene            <9619. .>13591
                     /gene="CG8420"
                     /map="85D2-85D4"
                     /db_xref="FLYBASE:FBan0008420"
                     /db_xref="FLYBASE:FBgn0037664"
     mRNA            join(9619. .10588,10978. .11126,12073. .12352,12809. .13132,
                     13366. .13591)
                     /gene="CG8420"
                     /product="CT24697"
                     /db_xref="FLYBASE:FBan0008420"
                     /db_xref="FLYBASE:FBgn0037664"
```

```
     CDS             join(9827..10588,10978..11126,12073..12352,12809..13132,
                     13366..13446)
                     /gene="CG8420"
                     /note="CG8420 gene product"
                     /codon_start=1
                     /protein_id="AAF54343.1"
                     /db_xref="GI:7299145"
                     /db_xref="FLYBASE:FBan0008420"
                     /db_xref="FLYBASE:FBgn0037664"
                     /translation="MYVRRALLLACLLCLQPLGPSMASEDESNPLLDMASMFFQEALS
                     NQNGGNNGGGGAGLAGVASLIGTFMQASGKSGGAGAGGGGAGGAMQILSGLGSLLSKS
                     QGGQSGGFDPSIIGNVLEMFTQGDDEEATPQQKRSNGGGSESGIGLDTILQVASAFMN
                     TQGNDKATHHHQQKRSTATEPESENGLMNLLPLVMQAVSSFAGPEGQSTQEKHKSHAW
                     VLPPFLEHIHVLWDHFSNSELADALYEKSGVNKIMKGFKGSDGKLDYDKLFESLNNQS
                     FRRRWIKSATLYLADWASYLANPEVYLRYFQTAQIMFNGLLKSQGYPKQTHFDPSRPG
                     ETISNLLDHVAKHHLNVKIDSRQYVKPAVGYAKELLKLGQARGLLQFNATEISDKLTD
                     TLNLEVIEPVLKVHRAYRYISKSPQCDRYVLCQLNAAALDQQEKQRQHDQYQQHHQPK
                     QSQLNRPTSASSLIAGVSPKIVKIGSMGAAIFISTETGTPFWTLFGVINAPYNCEAKY
                     PVDCNGFHEGEAKVTTEYIHNEL"
     gene            <14006..>15118
                     /gene="CG16733"
                     /map="85D2-85D4"
                     /db_xref="FLYBASE:FBan0016733"
                     /db_xref="FLYBASE:FBgn0037665"
     mRNA            join(14006..14358,14503..>15118)
                     /gene="CG16733"
                     /product="CT37237"
                     /db_xref="FLYBASE:FBan0016733"
                     /db_xref="FLYBASE:FBgn0037665"
     CDS             join(14024..14358,14503..15118)
                     /gene="CG16733"
                     /note="CG16733 gene product"
                     /codon_start=1
                     /protein_id="AAF54344.1"
                     /db_xref="GI:7299146"
                     /db_xref="FLYBASE:FBan0016733"
                     /db_xref="FLYBASE:FBgn0037665"
                     /translation="MQLIYRELEEDILRRTNAVFPVQNCFVEVLPDQFIIPRKYVELG
                     ESIRSLPVYQDDVWMVSYPRTGSTWAQEMVWLLGHQLDYVAAEQDLRLRSPLIELSAL
                     FSIDHHETVAQKFGNTVDLVRNLPRPRFARSHLPWPLLPEQFETVKPRIVYTARNPKD
                     LCVSYYHYFKLLHGMNGDFEQFVDLFLEGHTPMGSYWRHVLPFWKRSQDDNVLFIKYE
                     DMVKDLPSVVRRCARFLGVQSLLDVSTLQKLCDHLTFDKMRANKAVNLEKLLPESSSK
                     FIRNGKIGDWRNHMGNEMSERFDEWTERHMRGSGLNFDYV"
     gene            <16163..>16836
                     /gene="CG16735"
                     /map="85D2-85D4"
                     /db_xref="FLYBASE:FBan0016735"
                     /db_xref="FLYBASE:FBgn0037666"
     mRNA            join(<16163..16213,16307..16490,16637..>16836)
                     /gene="CG16735"
                     /product="CT37245"
                     /db_xref="FLYBASE:FBan0016735"
                     /db_xref="FLYBASE:FBgn0037666"
     CDS             join(16163..16213,16307..16490,16637..16836)

  Query Match             2.0%;  Score 42.2;  DB 3;  Length 224400;
  Best Local Similarity   49.3%;  Pred. No. 12;
  Matches  110;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy       1470 GGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGC 1529
Db      24013 GGCCGCTATCAGCCGCCTGCTTGCTGCCTCCCAGAAGTCCAGCGACTCGGTTCTGTTTGT 24072

Qy       1530 CAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
Db      24073 CCAGACCACGGAGAAGGACGTGGCAGCATCGCGTGCCAAGCGTGATACCATTGCCGCTTC 24132

Qy       1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
Db      24133 CACCACCAATCCCTACAACCTGGCCGTCTACTACGGCAGCGACTACCCGGTGATCTTCAA 24192

Qy       1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCG 1692
Db      24193 CATCATCCTGTGGTTCATGGTCGTCTTCGGACTGTCTCTGCTG 24235


RESULT 10
CNS01DBA
LOCUS       CNS01DBA                 720 bp    mRNA    linear   PLN 03-SEP-1999
DEFINITION  Botrytis cinerea strain T4 cDNA library under conditions of
            nitrogen deprivation.
ACCESSION   AL116606
VERSION     AL116606.1  GI:5831822
KEYWORDS    cDNA library; nitrogen deprivation.
SOURCE      Botryotinia fuckeliana.
  ORGANISM  Botryotinia fuckeliana
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
            Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
            78026 Versailles, France
REFERENCE   2  (bases 1 to 720)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
            CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     The cDNA library to be analyzed within the framework of this
            project was created using a Botrytis cinerea strain which was grown
            under conditions of nitrogen deprivation, which is the normal
            situation for B. cinerea during its development on its host plant.
            The library was produced in an oriented direction, in the pBSII
            vector.
FEATURES             Location/Qualifiers
     source          1..720
                     /organism="Botryotinia fuckeliana"
                     /strain="T4"
                     /db_xref="taxon:40559"
                     /note="Genoscope sequence ID : W45G011"
BASE COUNT      229 a    199 c    130 g    161 t      1 others
ORIGIN

  Query Match             1.9%;  Score 41.6;  DB 8;  Length 720;
  Best Local Similarity   57.8%;  Pred. No. 8.4;
  Matches   74;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy       1482 CGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGA 1541
Db        204 CGAGACTCTCAAGGCCAAATTCAAGGCTCTCCTCAAAGGAAAGAAATCCAAGAAGGCCGA 263

Qy       1542 GAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGC 1601
Db        264 GGCTGAACCAACTGCCGATGCAACAAAGACCGATGCAGCTGCTGCTCCAGCTACTGCAAC 323

Qy       1602 CGATGCGA 1609
Db        324 CGATGCTA 331


RESULT 11
RHMNTRBC
LOCUS       RHMNTRBC                1700 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  R.meliloti ntrC gene, complete cds, and ntrB gene, 3' end.
ACCESSION   M15810
VERSION     M15810.1  GI:152391
KEYWORDS    ntrB protein gene; ntrC protein gene.
SOURCE      R.meliloti DNA.
  ORGANISM  Sinorhizobium meliloti
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Sinorhizobium.
REFERENCE   1  (bases 1 to 1700)
  AUTHORS   Szeto,W.W., Nixon,B.T., Ronson,C.W. and Ausubel,F.M.
  TITLE     Identification and characterization of the Rhizobium meliloti ntrC
            gene: R. meliloti has separate regulatory pathways for activation
```

```
              of nitrogen fixation genes in free-living and symbiotic cells
   JOURNAL    J. Bacteriol. 169 (4), 1423-1432 (1987)
   MEDLINE    87165745
    PUBMED    2881918
FEATURES                Location/Qualifiers
     source             1. .1700
                        /organism="Sinorhizobium meliloti"
                        /db_xref="taxon:382"
     CDS                <1. .255
                        /note="ntrB"
                        /codon_start=1
                        /transl_table=11
                        /protein_id="AAA26345.1"
                        /db_xref="GI:152392"
                        /translation="EFCVHDNGPGVPPDLLPHLFDPFITTKTNGSGLGLALVAKIIGG
                        HGGIVECDSQHSRTTFRVLMPASKGLAADEETPMTKGTNG"
     CDS                252. .1700
                        /note="ntrC"
                        /codon_start=1
                        /transl_table=11
                        /protein_id="AAA26346.1"
                        /db_xref="GI:152393"
                        /translation="MTGATILVADDDAAIRTVLNQALSRAGYDVRITSNAATLWRWIA
                        AGDGDLVVTDVVMPDENAFDLLPRIKKARPDLPVLVMSAQNTFMTAIKASEKGAYDYL
                        PKPFDLTELIGIIGRALAEPKRRPSKLEDDSQDGMPLVGRSAAMQEIYRVLARLMQTD
                        LTLMITGESGTGKELVARALHDYGKRRNGPFVAINMAAIPRDLIESELFGHEKGAFTG
                        AQTRSTGRFEQAEGGTLFLDEIGDMPMDAQTRLLRVLQQGEYTTVGGRTPIRSDVRIV
                        AATNKDLKQSINQGLFREDLYYRLNVVPLRLPPLRDRAEDIPDLVRHFVQQAEKEGLD
                        VKRFDQEALELMKAHPWPGNVRELENLVRRLTALYPQDVITREIIENELRSEIPDSPI
                        EKAAARSGSLSISQAVEENMRQYFASFGDALPPSGLYDRVLAEMEYPLILAALTATRG
                        NQIKAADLLGLNRNTLRKKIRELGVSVYRSLA"
BASE COUNT      341 a    553 c    511 g    295 t
ORIGIN

  Query Match             1.9%;  Score 41;  DB 1;  Length 1700;
  Best Local Similarity   55.2%;  Pred. No. 14;
  Matches   80;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy       1654 GCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATG 1713
              ||||   |||  || || |||| ||  | ||| | || ||    |  || || ||||  ||
Db        681 GCCGCCATGCAGGAAATCTACCGCGTGCTTGCGCGGCTGATGCAGACCGACCTCACGCTG 740

Qy       1714 CGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTAT 1773
                 || | |  ||  |||  ||  ||  | ||||      | | ||||   | || |||||
Db        741 ATGATCACCGGCGAGTCCGGCACCGGCAAGGAACTCGTTGCCCGTGCACTGCACGACTAT 800

Qy       1774 CATACGACAGGCAAGGGCGTATTGG 1798
                 | |    | || |||    || |
Db        801 GGCAAGCGGAGAAATGGCCCCTTCG 825


RESULT 12
SME591787
LOCUS       SME591787              329100 bp    DNA     linear   BCT 05-JUL-2002
DEFINITION  Sinorhizobium meliloti 1021 complete chromosome; segment 6/12.
ACCESSION   AL591787 AL591688
VERSION     AL591787.1  GI:15074266
KEYWORDS    .
SOURCE      Sinorhizobium meliloti.
  ORGANISM  Sinorhizobium meliloti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Sinorhizobium.
REFERENCE   1  (bases 1 to 329100)
  AUTHORS   Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
            Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
            Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
            Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
            Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
  TITLE     Analysis of the chromosome sequence of the legume symbiont
            Sinorhizobium meliloti strain 1021
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
  MEDLINE   21396507
   PUBMED   11481430
REFERENCE   2  (bases 1 to 329100)
  AUTHORS   Gouzy,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
            EU Consortium
COMMENT     MELILO EU Consortium:
            Laboratoire de Biologie Moleculaire des Relations
            Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
            France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
            Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
            France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
            Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
            D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
            Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
            B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
            des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
            B-5030 Gembloux, Belgium.  E-mail:Jerome.Gouzy@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/meliloti.html.
FEATURES             Location/Qualifiers
     source          1. .329100
                     /organism="Sinorhizobium meliloti"
                     /strain="1021"
                     /db_xref="taxon:382"
     gene            152. .227
                     /gene="tRNA-THR_TGT"
     tRNA            152. .227
                     /gene="tRNA-THR_TGT"
                     /product="tRNA-Thr"
                     /note="codon recognized: ACA; predicted by tRNAscan-SE"
                     /evidence=not_experimental
     gene            complement(373. .798)
                     /gene="SMc04435"
     CDS             complement(373. .798)
                     /gene="SMc04435"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Homology"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="CONSERVED HYPOTHETICAL PROTEIN"
                     /protein_id="CAC45913.1"
                     /db_xref="GI:15074267"
                     /db_xref="SPTREMBL:Q92QI9"
                     /translation="MNNALVPEFAVSDWQASKEFYCNILGFSCDYERPEEGFCYLSLD
                     GAELMIDQIGTGRTFADGHLPRSYPFGRGLNVQIRVASADALVRALTDRGIALYLPLE
                     ERWYRRGDEEVGNRQFVVADPDGYLLRFYEPLGRRPRTS"
     gene            complement(914. .1855)
                     /gene="SMc01330"
     CDS             complement(914. .1855)
                     /gene="SMc01330"
                     /function="miscellaneous; not classified regulator"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
                     /protein_id="CAC45914.1"
                     /db_xref="GI:15074268"
                     /db_xref="SPTREMBL:Q92QI8"
                     /translation="MQDLNDIALFAAVVKHNGFSAAARDLNVPKSKLSKHVARLEEQL
                     GVRLLERSTRKLRMTEVGRIFYEHAQGLLDGVAAAEAEIAAVRAEPTGVVRLACPIGF
                     TPMLADILPAFHRRYPGVRLLITASNRRIDLIEERIDVALRARDQLDTDSQLIVRKFG
                     EVRQRLAASPTLLARLGEITTDNLSQMPTLSMNEQHPNDVWRLVHATGGAIEIAHRPV
                     VGCSDFLILERAAIEGMGIALLPDHICERAFRTGALVPVLPEWTSGNVMVHLVFPSRH
                     GLLPATRALIDFLAENLIKALERCREVDPRPAASFEI"
     gene            2063. .2683
                     /gene="acpD OR SMc01329"
```

```
     CDS             2063..2683
                     /gene="acpD OR SMc01329"
                     /function="small molecule metabolism; fatty acid
                     biosynthesis"
                     /note="Product confidence : probable
                     Gene name confidence : putative
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE ACYL CARRIER PROTEIN PHOSPHODIESTERASE"
                     /protein_id="CAC45915.1"
                     /db_xref="GI:15074269"
                     /db_xref="SPTREMBL:Q92QI7"
                     /translation="MKILHIDSGILGEHSVSRRLTSAIVSQLKADRPDAEITYRDLAS
                     ERVPHLTGAQIMAPADLEGVDALLAADVRIGRQMLEEFLAADTVVVGAPMYNFSIPSQ
                     LKAWIDRLAVAGKTFRYTEAGPEGLAKGKKLIVASTRGGHYSVAPASAMDHQETYLRS
                     VFGFFGITDIEFIRAEGLNLGPDQKQFAIAEAEKTIAEGNVLKLAS"
     gene            2792..3076
                     /gene="SMc01328"
     CDS             2792..3076
                     /gene="SMc01328"
                     /function="miscellaneous; hypothetical/partial homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
                     /protein_id="CAC45916.1"
                     /db_xref="GI:15074270"
                     /db_xref="SPTREMBL:Q92KE9"
                     /translation="MKCVRALRFLALSLAAVALVAGSVSASAPGGRCTCRHRDGAKYE
                     LGQTACIRIGDISYLARCEMNLNVTTWKRLQDGCPTAGLAVTIVPSAPIP"
     gene            complement(3066..3941)
                     /gene="SMc01327"
     CDS             complement(3066..3941)
                     /gene="SMc01327"
                     /EC_number="2.1.1.-"
                     /function="macromolecule metabolism; macromolecule
                     synthesis, modification; rna synthesis, modification , dna
                     transcription"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PUTATIVE TRNA/RRNA METHYLTRANSFERASE PROTEIN"
                     /protein_id="CAC45917.1"
                     /db_xref="GI:15074271"
                     /db_xref="SPTREMBL:Q92QI6"
                     /translation="MSKDKTGDKSARDTHYANLRRAHRDARRERGEIPTPRQDKRAKL
                     SADWKPPALAPDQVLLYGLHTVRAALDNPERRIIRLSVTQNAAARLDLGDLTALPFPT
                     ETVMPQELDRILGPDAIHQGVMLETQPLPHRKLEALRDCPLILVLDQVTDPHNVGAIM
                     RSAVAFDAGALITTMRHSPTESGVLAKSASGALELIPYIQITNLADTLEELHELGFMT
                     IGLDSEGPQPLEGTFAGDKIALVLGSEGKGLRQKTRQTVKALARLDMPGAIKSLNVSN
                     AAAIAMYAARHHLKG"
     gene            4158..4242
                     /gene="tRNA-TYR_GTA"
     tRNA            4158..4242
                     /gene="tRNA-TYR_GTA"
                     /product="tRNA-Tyr"
                     /note="codon recognized: UAC; predicted by tRNAscan-SE"
                     /evidence=not_experimental
     gene            4266..4339
                     /gene="tRNA-GLY_TCC"
     tRNA            4266..4339
                     /gene="tRNA-GLY_TCC"
                     /product="tRNA-Gly"
                     /note="codon recognized: GGA; predicted by tRNAscan-SE"
                     /evidence=not_experimental
     gene            4602..5777
                     /gene="tufB OR SMc01326"
     CDS             4602..5777
                     /gene="tufB OR SMc01326"
                     /function="macromolecule metabolism; macromolecule
                     synthesis, modification; proteins and peptides -
                     translation and modification"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE ELONGATION FACTOR TU PROTEIN"
                     /protein_id="CAC45918.1"
                     /db_xref="GI:15074272"
                     /db_xref="SPTREMBL:Q925Y6"
                     /translation="MAKSKFERNKPHVNIGTIGHVDHGKTSLTAAITKYFGEFKAYDQ
                     IDAAPEEKARGITISTAHVEYETPNRHYAHVDCPGHADYVKNMITGAAQMDGAILVVS
                     AADGPMPQTREHILLARQVGVPAIVVFLNKVDQVDDAELLELVELEVRELLSSYEFPG
                     DDIPIVKGSALAALEDSDKKIGEDAIRELMAAVDAYIPTPERPIDQPFLMPIEDVFSI
                     SGRGTVVTGRVERGIVKVGEEIEIVGIRPTTKTTCTGVEMFRKLLDQGQAGDNIGALL
                     RGVDRNGVERGQILCKPGSVKPHRKFKAEAYILTKEEGGRHTPFFTNYRPQFYFRTTD
                     VTGIVTLPEGTEMVMPGDNVTVDVELIVPIAMEEKLRFAIREGGRTVGAGIVASIVE"
     gene            6125..6790
                     /gene="SMc01325"
     CDS             6125..6790

  Query Match             1.9%;  Score 41;  DB 1;  Length 329100;
  Best Local Similarity   55.2%;  Pred. No. 26;
  Matches   80;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy       1654 GCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATG 1713
              ||||   |||   || || |||| ||  | |||  | || ||     |   || || ||||  ||
Db     117496 GCCGCCATGCAGGAAATCTACCGCGTGCTTGCGCGGCTGATGCAGACCGACCTCACGCTG 117555

Qy       1714 CGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTAT 1773
                 || | |  ||  |||  ||  ||  | ||||     | | ||||   | || |||||
Db     117556 ATGATCACCGGCGAGTCCGGCACCGGCAAGGAACTCGTTGCCCGTGCACTGCACGACTAT 117615

Qy       1774 CATACGACAGGCAAGGGCGTATTGG 1798
                 | |    | || |||   || |
Db     117616 GGCAAGCGGAGAAATGGCCCCTTCG 117640


RESULT 13
AF429315
LOCUS       AF429315               125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
```

```
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1..125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581..35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     gene            complement(<36507..>36887)
                     /gene="JPH3"
                     /note="JP3"
     mRNA            complement(<36507..>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     CDS             complement(<36507..36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN

  Query Match             1.9%;  Score 40.4;  DB 9;  Length 125020;
  Best Local Similarity   9.7%;  Pred. No. 33;
  Matches   79;  Conservative  349;  Mismatches  379;  Indels    6;  Gaps    2;

Qy         27 TCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGGGTAGCCAAAGCCGACAAAGGCATGTG 86
              :|::::   :: :: :   ::::: :|:  : :::  :  |   :||  |  :    :|     :
Db      17023 KCYMSSYWTMSASYMSRRSYWKRSMKCWSRMCSRSRSSKCKSRGCGGSMGKGKGGKSYGR 17082

Qy         87 GCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCT 146
              : :        :: | ::: : :: :::   : ::   :  :     |   ::||:|: :|     |
Db      17083 KTKRKSRGKMKGAKWYMYYRRSMRKMKMYSSKGMYCMYCWCWGRRGCYCSCMTSRSAMCC 17142

Qy        147 CCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTT 206
              ::|   :  :     ::|:: :::: :::   : :|:::   :     |:     |:|       || ||
Db      17143 SYCYAKCKSMCYSCYYGMSMKGYYYACSYRGSMSSKYCMRGSTYSTSCGCCCTTTTTCCC 17202

Qy        207 CGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAACCACCA 266
              |         ||          | | |   | :: ::: ::   |   :|:    ::        | | :
Db      17203 CCNANTGGG----GAAGCTTTTNCNKKTYSYYRKRNGCAMCKYNNYYNNSWRSSCRAGMS 17258

Qy        267 CTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTATCTGCGCGATGG 326
              ||::::   ::   :: :: :::    ::: : ::   ::::    ::: ::: :   ::| :   :
Db      17259 CTKYKSSMTMSMASYCWCMSMYYCSMRSMASRGMSWSSYMKMMKSSWRMSCYMKCCWCMK 17318

Qy        327 TTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCT 386
                 :|      : :: ::       ::: :::||:       : ::::: :: :::    :    :    ::
Db      17319 MCYCMMRSMRSRGMSYYMYASWKSSSRGCYCTRCYWCMSSKSCYKSYYMMMRSKRMKGMK 17378

Qy        387 GCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGA 446
                ::|: ::: : :: :  :   ::   ::: : ::      :||||  :: ||   ::::|::: ::
Db      17379 WSRCWGSMWGSASRSSCYKCYK--SMRCSMMSSKCYRCAGCMMMKGGYMRYMRCWSMKRR 17436

Qy        447 GATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGC 506
                : ::| ::         : : | : |:: ::         :::::   |  :   : |::     :     ::
Db      17437 WWGKSAMYRMRRMWKGRGAMMMCMKCYSRMSRCMMMMKSYCASCRSCAMMSGTYMKCASY 17496

Qy        507 AGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCAT 566
              : :::       |      :::   ::     : :    :::|:::  : ::: :      |::: :   |||::

Db      17497 MCMSSYWCTCWCTSYMSYYRCTCWCKGWSSYTKSKSWSSSMSSYKKGRRKSYSMCCTSRG 17556

Qy        567 CGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAA 626
                   |::    :        ::: | ::        :    :          :           :    ::::: | :: ::
Db      17557 AMSCWRRCCYMRGASSMRAGSMSRRRAKGRSWGGRSKWWMTGGMWRSKYYYYCTGRRMMM 17616

Qy        627 GTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATT 686
                 : |   ::: :    :::::|:   ::        |   :   :  :  :|:: :: :: ::::: :
Db      17617 TYMCCCWRRRSYMYYRSMSAMGMRKSSWSGMRMMGSASSRRCKSASRSSWCSRRMKGMRG 17676

Qy        687 CCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTA 746
              :| ::: :    :  ::: ::          : ::: ::: :|   ::                :    :
Db      17677 SCWSSKMWGGSRSRSASSKCKGSRGMRRRSKRSSKYRKRGRGKKRSMTKSKGSKGKSKCW 17736

Qy        747 TAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCAT 806
                :: |:: :    :  :   ::|   :::     |:      :::::   ||    :::      :        :| :
Db      17737 KRSGSMTSSCYYYSASSCMWMMSSKSCMCCCMMMKRRCACCYSSMSCTSMYRCCWGGMKS 17796

Qy        807 GACCATCGGTTTCCCGGGCAGTACGGATCGCTA 839
                   | |    ::  |  : ::: :  ::: ::|:::::
Db      17797 YSTCGTTCKKCTGYKKSRTWYWWMYSWTSKMYM 17829


RESULT 14
AF051784
LOCUS       AF051784                4314 bp    mRNA    linear   VRT 18-JUL-1998
DEFINITION  Xenopus laevis 14S cohesin SMC1 subunit mRNA, complete cds.
ACCESSION   AF051784
VERSION     AF051784.1  GI:3328230
KEYWORDS    .
SOURCE      Xenopus laevis.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 4314)
  AUTHORS   Losada,A., Hirano,M. and Hirano,T.
  TITLE     Identification of Xenopus SMC protein complexes required for sister
            chromatid cohesion
  JOURNAL   Genes Dev. 12 (13), 1986-1997 (1998)
  MEDLINE   98315077
   PUBMED   9649503
REFERENCE   2  (bases 1 to 4314)
  AUTHORS   Losada,A., Hirano,M. and Hirano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-1998) Cold Spring Harbor Laboratory, 1 Bungtown
            Road, Cold Spring Harbor, NY 11724, USA
FEATURES             Location/Qualifiers
     source          1..4314
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     CDS             158..3856
                     /note="member of structural maintenance of chromosomes
                     family of ATPases; SMC protein"
                     /codon_start=1
                     /product="14S cohesin SMC1 subunit"
                     /protein_id="AAC26807.1"
                     /db_xref="GI:3328231"
                     /translation="MGFLKLIEIENFKSYKGRQIIGPFHRFTAIIGPNGSGKSNLMDA
                     ISFVLGEKTSNLRVKTLRDLIHGAPVGKPAANRAFVSMVYSEDSGEEKVFSRVIVGGS
                     SEYKINNKVVQLSEYSDSLEKLGILIKARNFLVFQGAVESIAMKNPKERTALFEEISR
                     SGELAQEYDKRKKEMVKAEEDTQFNYHRKKNIAAERKEAKQEKEEAERYQRLKDEVAR
                     AQIQLQLFKLYHNESEIEKLNKELSVKNKGIEKDKKHMDKVEEELKDKKKELGKMMRE
                     QQAIEKEIKEKDAELNQKLPQYIKAKENPSHKIKKFRAAKKSLQNAQKQYKKRKADMD
                     ELEKEMLSVEKARQEFEERMEEESQSQGRDLTLEENQVKKYHRLKEEASKRAATLAQE
                     LEKFNRDQKADQDRLDLEERKKVETEAKIKQKLRELEENQKRIEKLEEYIATSKQSLE
                     EQKNLEETLTEEVEMAKRRIDEINSELNQVMEQLGDARIDRQESSRQQRKAEIMESIK
                     RLYPGSVYGRLIDLCQPTQKKYQIAVTKVLGKNMDAIIVDSEKTGRDCIQYIKEQRGE
                     PETFLPLDYLEVKPTDERLRELKGAKLVIDVIRYEPPHIKKALQYACGNALVCDNVED
                     ARRIAFGGHQRHKTVALDGTLFQKSGVISGGASDLKAKARRWDEKAVDKLKEKKERLT
                     EELKEQMKAKRKEAELRQVQSQAHGLQMRLKYSQSDLEQTKTRHLAMNMQEKSKLESE
                     LANFSPRINDIKRIIQSRDREMKDLKEKMNQVEDEVFEEFCREIGVRNIREFEEEKVK
                     RQNEIAKKRLEFENQKTRLGIQLDYEKNQLKEDQGKVQTWEQSVKKDDNEIEKLKKEE
```

```
                QRHMKIIDETMAQLQDLKNQHLAKKSEVNDKNHLMEDIRKKLGSANKEVTHLQKEVTA
                IETKLEQKRSDRHNLLQACKMSDIKLPLSKGTMDDISQEEGSSQGEESASSSQRSSTV
                YAKEALIEIDYSDLSEDLKDAVADDDIKQEMSALHQKINEQQSILQRISAPNMKAMEK
                LESVRDKFQETSDEFEAARKRAKKAKQAFEQTKKERFDRFNACFESVATNIDEIYKAL
                SRNSSAQAFLGPENPEEPYLDGINYNCVAPGKRFRPMDNLSGGEKTVAALALLFAIHS
                YKPSPFFVLDEIDAALDNTNIGKVANYIKEQSMSNFQAIVISLKEEFYTKAESLIGVY
                PEQGDCVISKVLTFDLTKYPDANPNPND"
BASE COUNT     1408 a    821 c    1141 g     944 t
ORIGIN

  Query Match             1.9%;  Score 40;  DB 5;  Length 4314;
  Best Local Similarity   58.3%;  Pred. No. 28;
  Matches   70;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        399 GAAGGTAACGGACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCT 458
              ||||   | ||||||||| || | | |||||||||  |   |    | ||| ||| |    |
Db        916 GAAGCATATGGACAAGGTGGAGGAAGAGCTCAAGGACAAGAAGAAAGAGTTGGGGAAGAT 975

Qy        459 GCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCA 518
                 |    |  ||   ||| |||        |   ||   |   |||| |    |   |||||||     ||| ||
Db        976 GATGAGAGAGCAGCAGGCCATTGAGAAAGAGATCAAAGAGAAGGATGCAGAACTGAATCA 1035


RESULT 15
AC121850
LOCUS       AC121850              202306 bp    DNA     linear   HTG 21-MAY-2002
DEFINITION  Mus musculus chromosome UNK clone RP24-92E18, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC121850
VERSION     AC121850.1  GI:21039949
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 202306)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 202306)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ---------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BB0092E18
            ---------------------------------------------------
            -------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 197589 bases at least Q40
            Consensus quality: 198358 bases at least Q30
            Consensus quality: 198775 bases at least Q20
            Insert size: 172000; agarose-fp
            Insert size: 201306; sum-of-contigs
            Quality coverage: 21.02 in Q20 bases; agarose-fp
            Quality coverage: 11.20 in Q20 bases; sum-of-contigs
            ----------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
```

```
            * as soon as it is available and the accession number will
            * be preserved.
            *          1      1106: contig of 1106 bp in length
            *       1107      1206: gap of unknown length
            *       1207      2283: contig of 1077 bp in length
            *       2284      2383: gap of unknown length
            *       2384      4051: contig of 1668 bp in length
            *       4052      4151: gap of unknown length
            *       4152      5430: contig of 1279 bp in length
            *       5431      5530: gap of unknown length
            *       5531      6547: contig of 1017 bp in length
            *       6548      6647: gap of unknown length
            *       6648      8792: contig of 2145 bp in length
            *       8793      8892: gap of unknown length
            *       8893     21575: contig of 12683 bp in length
            *      21576     21675: gap of unknown length
            *      21676     43695: contig of 22020 bp in length
            *      43696     43795: gap of unknown length
            *      43796     95454: contig of 51659 bp in length
            *      95455     95554: gap of unknown length
            *      95555    145463: contig of 49909 bp in length
            *     145464    145563: gap of unknown length
            *     145564    202306: contig of 56743 bp in length.
FEATURES             Location/Qualifiers
     source          1..202306
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP24-92E18"
     misc_feature    1..1106
                     /note="assembly_name:Contig25"
     misc_feature    1207..2283
                     /note="assembly_name:Contig28"
     misc_feature    2384..4051
                     /note="assembly_name:Contig32"
     misc_feature    4152..5430
                     /note="assembly_name:Contig34"
     misc_feature    5531..6547
                     /note="assembly_name:Contig35"
     misc_feature    6648..8792
                     /note="assembly_name:Contig36
                     clone_end:SP6
                     vector_side:right"
     misc_feature    8893..21575
                     /note="assembly_name:Contig37"
     misc_feature    21676..43695
                     /note="assembly_name:Contig38"
     misc_feature    43796..95454
                     /note="assembly_name:Contig39"
     misc_feature    95555..145463
                     /note="assembly_name:Contig40"
     misc_feature    145564..202306
                     /note="assembly_name:Contig41"
BASE COUNT    55798 a  44381 c  43982 g  57115 t   1030 others
ORIGIN

  Query Match             1.9%;  Score 40;  DB 2;  Length 202306;
  Best Local Similarity   52.4%;  Pred. No. 45;
  Matches   88;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy       1444 GCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTC 1503
              | |  ||||    |  ||||| ||| ||  || | |||   |      |        |||||  |
Db     147941 GTACCCTTCAACTCTGACAATAGTATGTATCTTCATATGGTGGTGGGGTTGGCCATCTTA 148000

Qy       1504 AAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTT 1563
              |    |      ||  ||  || ||  |    |||| | | || |  |      ||||| | |||
Db     148001 ATCCCATGAGATGAGCTAACGTAAGAATAGGCAAACAAGGAGAACAGAGCAGTGGGGCTG 148060

Qy       1564 TCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATG 1611
              || | ||||      ||||| ||| | |      ||||   | |  ||  ||
Db     148061 TCAAGGAGCCCTTTAGCTTCTGTTGGAATGATTCCATCTGGAGCTCTG 148108
```

Search completed: May 23, 2003, 05:02:07
Job time : 5341 secs

```
OM nucleic - nucleic search, using sw model

Run on:         May 23, 2003, 02:20:11 ; Search time 327 Seconds
                                         (without alignments)
                                         14730.961 Million cell updates/sec

Title:          US-10-008-355-1
Perfect score:  2139
Sequence:       1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters:      4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_101002:*
                 1:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                 2:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                 3:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                 4:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                 5:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                 6:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                 7:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                 8:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                 9:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                 10:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                 11:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                 12:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                 13:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                 14:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                 15:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                 16:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                 17:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                 18:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                 19:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
                 20:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                 21:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                 22:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                 23:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                 24:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2139	100.0	2139	24	AAL43635	Porphyromonas ging
	2	42.2	2.0	1147	23	ABL14089	Drosophila melanog
	3	42.2	2.0	3632	23	ABL14088	Drosophila melanog
	4	40.6	1.9	2868	23	ABL21554	Drosophila melanog
	5	38.8	1.8	4013	23	ABL18653	Drosophila melanog
c	6	38.6	1.8	3621	22	AAH48797	P. pantotrophus GB
c	7	38.6	1.8	4848	22	AAH48798	P. pantotrophus GB
	8	37.8	1.8	10757	23	ABL21478	Drosophila melanog
c	9	37.4	1.7	9845	23	ABL18652	Drosophila melanog
	10	37	1.7	786	22	AAC68978	Merozoite surface
	11	37	1.7	1077	22	AAC68976	Merozoite surface
c	12	36.8	1.7	810	22	AAH01671	Fonsecaea pedrosoi
c	13	36.8	1.7	349980	22	AAF86431	Pyrococcus abyssi
	14	36.4	1.7	1090	21	AAC66137	Human kidney disea
	15	36.4	1.7	1292	22	AAI58976	Human polynucleoti
	16	36.4	1.7	1649	21	AAZ35391	Human kidney-speci
c	17	36.4	1.7	1768	22	AAI60762	Human polynucleoti
	18	36.2	1.7	345	22	AAH67046	C glutamicum codin
	19	36.2	1.7	536	21	AAA43983	Mouse secreted exp
	20	36.2	1.7	2280	17	AAT36127	Mouse neuropeptide
	21	36.2	1.7	2281	24	AAI72369	NPY6-R gene. Mus
	22	36.2	1.7	349980	22	AAH68530	C glutamicum codin
	23	36	1.7	292	24	ABL71482	Corn tassel-derive
	24	35.6	1.7	6174	22	AAK81576	Human immune/haema
	25	35.6	1.7	8617	22	AAK81577	Human immune/haema
	26	35.6	1.7	32313	22	AAK81578	Human immune/haema
	27	35.2	1.6	449	20	AAV68139	DNA encoding a hum
	28	35.2	1.6	4590	22	AAH24065	Yeast AOD9604-asso
	29	35	1.6	1291	24	ABK65331	Arabidopsis cDNA e
c	30	35	1.6	1910	23	AAS85131	DNA encoding novel
c	31	34.8	1.6	632	21	AAF08951	Fusarium venenatum
c	32	34.8	1.6	1368	24	ABN68375	Streptococcus poly
c	33	34.8	1.6	5059	20	AAX84332	Stealth virus nucl
	34	34.6	1.6	639	21	AAF13795	Aspergillus oryzae
c	35	34.6	1.6	672	24	ABQ90235	M. capsulatus gene
	36	34.6	1.6	975	23	ABL16623	Drosophila melanog
	37	34.6	1.6	2975	23	ABL16622	Drosophila melanog
	38	34.6	1.6	5133	23	ABL09568	Drosophila melanog
c	39	34.4	1.6	1163020	24	ABQ67197	Listeria innocua c
	40	34.2	1.6	3059	8	AAN70566	Sequence encoding
c	41	34	1.6	2759	23	ABL06895	Drosophila melanog
	42	34	1.6	3972	24	AAS97214	Neisseria meningit
c	43	34	1.6	5837	12	AAQ14461	Polymerase gene.
c	44	34	1.6	5837	14	AAQ43966	Thermococcus litor
c	45	34	1.6	5837	14	AAQ43515	T. litoralis DNA p

ALIGNMENTS

```
RESULT 1
AAL43635
ID   AAL43635 standard; DNA; 2139 BP.
XX
AC   AAL43635;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) coding sequence.
XX
KW   Dipeptidylpeptidase-7; DPP-7; gene; ds; enzyme; amidolytic cleavage;
KW   DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW   periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2139
FT                   /*tag= a
FT                   /product= "Porphyromonas gingivalis DPP-7"
XX
PN   WO200238742-A2.
XX
PD   16-MAY-2002.
XX
PF   08-NOV-2001; 2001WO-US46782.
XX
PR   08-NOV-2000; 2000US-246827P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,  Potempa JS,  Banbula A,  Bugno M;
```

```
XX
DR   WPI; 2002-490075/52.
DR   P-PSDB; AAO15205.
XX
PT   Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT   the dipeptidylpeptidase for protecting an animal from periodontal
PT   disease caused by Porphyromonas gingivalis  -
XX
PS   Claim 11; Fig 4; 65pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC   enzyme has amidolytic activity for cleavage of a peptide bond between the
CC   second and third amino acids from the N-terminal end of a target peptide.
CC   The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC   substituent on the alpha-carbon atom of the second amino acid from the
CC   N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC   useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC   reducing the growth of a bacterium and protecting an animal from a
CC   periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC   or periodontitis). The present DNA sequence encodes the Porphyromonas
CC   gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the invention.
XX
SQ   Sequence 2139 BP; 544 A; 543 C; 565 G; 487 T; 0 other;

  Query Match             100.0%;  Score 2139;  DB 24;  Length 2139;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2139;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60

Qy       61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120

Qy      121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180

Qy      181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240

Qy      241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300

Qy      301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360

Qy      361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420

Qy      421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480

Qy      481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540

Qy      541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600

Qy      601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660

Qy      661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720

Qy      721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780

Qy      781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840

Qy      841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900

Qy      901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960

Qy      961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020

Qy     1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080

Qy     1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140

Qy     1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200

Qy     1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260

Qy     1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320

Qy     1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380

Qy     1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440

Qy     1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500

Qy     1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560

Qy     1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620

Qy     1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680

Qy     1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740

Qy     1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800

Qy   1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860

Qy   1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920

Qy   1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980

Qy   1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040

Qy   2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100

Qy   2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 2139
          |||||||||||||||||||||||||||||||||||||||
Db   2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 2139
```

```
RESULT 2
ABL14089
ID   ABL14089 standard; cDNA; 1147 BP.
XX
AC   ABL14089;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36749.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69986.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 36749; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1147 BP; 257 A; 335 C; 303 G; 252 T; 0 other;

  Query Match             2.0%;  Score 42.2;  DB 23;  Length 1147;
  Best Local Similarity   49.3%;  Pred. No. 0.11;
  Matches  110;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy   1470 GGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGC 1529
          ||   ||   |||  |  | |   ||||   |  |||||||  |||  ||    |||||
Db    711 GGCCGCTATCAGCCGCCTGCTTGCTGCCTCCCAGAAGTCCAGCGACTCGGTTCTGTTTGT 770

Qy   1530 CAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
          | || | |  ||||| ||   ||||| |  || | ||||| | |  |    ||| |||
Db    771 CCAGACCACGGAGAAGGACGTGGCAGCATCGCGTGCCAAGCGTGATACCATTGCCGCTTC 830

Qy   1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
          | | |   |  || |     |||||     ||| ||    ||  |      |    |
Db    831 CACCACCAATCCCTACAACCTGGCCGTCTACTACGGCAGCGACTACCCGGTGATCTTCAA 890

Qy   1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCG 1692
          | |   |   | |     | |   |  |||||   |||||| |
Db    891 CATCATCCTGTGGTTCATGGTCGTCTTCGGACTGTCTCTGCTG 933


RESULT 3
ABL14088
ID   ABL14088 standard; cDNA; 3632 BP.
XX
AC   ABL14088;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36746.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69985.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 36746; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
```

```
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3632 BP; 1062 A; 799 C; 773 G; 998 T; 0 other;

  Query Match             2.0%;  Score 42.2;  DB 23;  Length 3632;
  Best Local Similarity   49.3%;  Pred. No. 0.2;
  Matches  110;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy       1470 GGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGC 1529
                 ||   ||   |||  |  | |   ||||   |  |||||||  |||  ||     |||||
Db       2196 GGCCGCTATCAGCCGCCTGCTTGCTGCCTCCCAGAAGTCCAGCGACTCGGTTCTGTTTGT 2255

Qy       1530 CAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
              | || | |  ||||| ||   ||||| |  || |  ||||| |  |   |     ||| |||
Db       2256 CCAGACCACGGAGAAGGACGTGGCAGCATCGCGTGCCAAGCGTGATACCATTGCCGCTTC 2315

Qy       1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
              | | |    |   | ||  |      |||||      ||| |  ||  ||   |       |
Db       2316 CACCACCAATCCCTACAACCTGGCCGTCTACTACGGCAGCGACTACCCGGTGATCTTCAA 2375

Qy       1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCG 1692
              | |    |    | |       | |    |   |||||   |||||| |
Db       2376 CATCATCCTGTGGTTCATGGTCGTCTTCGGACTGTCTCTGCTG 2418


RESULT 4
ABL21554
ID   ABL21554 standard; DNA; 2868 BP.
XX
AC   ABL21554;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 16135.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD, .Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 16135; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2868 BP; 737 A; 657 C; 624 G; 850 T; 0 other;

  Query Match             1.9%;  Score 40.6;  DB 23;  Length 2868;
  Best Local Similarity   50.3%;  Pred. No. 0.53;
  Matches  100;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy       1494 TGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGC 1553
              ||||    |  |||||||   |||  ||        ||||| | || |  |  ||||| ||   |||
Db         10 TGCCTCCCAGAAGTCCAGCGACTCGGTTCTGTTTGTCCAGACCACGGAGAAGGACGTGGC 69

Qy       1554 AGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGC 1613
              || |  || |  ||||| | |   |      ||| |||   | | |    |   || |      ||||
Db         70 AGCATCGCGTGCCAAGCGTGATACCATTGCCGCTTCCACCACCAATCCCTACAACCTGGC 129

Qy       1614 CAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTA 1673
              |       ||| | ||  ||  |          | .   |    | |    |    | |       | |
Db        130 CGTCTACTACGGCAGCGACTACCCGGTGATCTTCAACATCATCCTGTGGTTCATGGTCGT 189

Qy       1674 CCCCGGACGTGCTCTGCCG 1692
              |   |||||    |||||| |
Db        190 CTTCGGACTGTCTCTGCTG 208


RESULT 5
ABL18653
ID   ABL18653 standard; DNA; 4013 BP.
XX
AC   ABL18653;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 7432.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 7432; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4013 BP; 968 A; 1026 C; 1086 G; 933 T; 0 other;
```

```
  Query Match             1.8%;  Score 38.8;  DB 23;  Length 4013;
  Best Local Similarity   47.2%;  Pred. No. 2.1;
  Matches  118;  Conservative    0;  Mismatches  132;  Indels    0;  Gaps    0;

Qy    1294 GACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGAT 1353
           |||||  | |  || |  |||  | | | | | |||| | | |  |  |   |||
Db     222 GACAATAATACGGAGTGGCTCATTTGGATGGTCGGTAATATTCCGGGCTGCGATGTCGCC 281

Qy    1354 ATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGAC 1413
           || |  |     || | | |  |  |||||| |  |     ||   |    | |   ||| |
Db     282 ATGGGTCAGACGCTTGTGGCCTACGACAACCGGCGGACCATCCACGGCAGTAACATCCAT 341

Qy    1414 AAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTT 1473
                |  ||       | |   ||| |||||   ||  |||  |||||| ||| | | |||
Db     342 CGAATCGTCTTCCTGGCCTTCAAGCAGTATCTGGAGCTCGATTTCGACGAGACTTTCGTT 401

Qy    1474 CCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAG 1533
           ||  |  ||||  |        ||| |  |   |     ||  |   |    |   |||  |
Db     402 CCCGAGGGCGAGGAAAAGGATGACGAGGACTTCACCTCGTCCGCTGGGTACGATGCAGAT 461

Qy    1534 GCTATCGAGA 1543
           | |   ||||
Db     462 GATGGGGAGA 471

RESULT 6
AAH48797/c
ID   AAH48797 standard; DNA; 3621 BP.
XX
AC   AAH48797;
XX
DT   05-NOV-2001  (first entry)
XX
DE   P. pantotrophus GB17 DNA encoding ORF2, ORF3 and soxXYZA.
XX
KW   Sulphur oxidation; sox; ORF1; ORF2; ORF3; soxX; soxY; soxZ; soxA;
KW   reduced sulphur; thiosulphate; transgenic plant; transgenic animal; ds.
XX
OS   Paracoccus pantotrophus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..738
FT                   /*tag= a
FT                   /transl_except= (pos:643..645,aa:Leu)
FT                   /product= "ORF2"
FT   CDS             752..1312
FT                   /*tag= b
FT                   /product= "ORF3"
FT   CDS             1427..1900
FT                   /*tag= c
FT                   /product= "soxX"
FT   CDS             1941..2363
FT                   /*tag= d
FT                   /product= "soxY"
FT   CDS             2385..2714
FT                   /*tag= e
FT                   /product= "soxZ"
FT   CDS             2749..3621
FT                   /*tag= f
FT                   /product= "soxA"
XX
PN   DE10006201-A1.
XX
PD   16-AUG-2001.
XX
PF   11-FEB-2000; 2000DE-1006201.
XX
PR   11-FEB-2000; 2000DE-1006201.
XX
PA   (FRIE/) FRIEDRICH C.
XX
PI   Friedrich C;
XX
DR   WPI; 2001-497622/55.
DR   P-PSDB; AAB86529, AAB86530, AAB86531, AAB86532, AAB86533, AAB86534.
XX
PT   New genes involved in oxidation of reduced sulphur compounds,
PT   particularly conversion of thiosulphate to sulphate, and related
PT   proteins, from Paracoccus pantotrophus -
XX
PS   Claim 1; Page 10-14; 22pp; German.
XX
CC   This invention describes novel genes (I) of the sox (sulphur oxidation)
CC   region designated ORF1, ORF2, ORF3 (Open Reading Frame), and soxXYZA
CC   from Paracoccus pantotrophus GB17. The proteins (II) encoded by (I) are
CC   used for oxidation of reduced sulphur compounds in cellular or cell-free
CC   systems, particularly of thiosulphate to sulphate. (I) is also used for
CC   expression of the proteins and to prepare transgenic plants or animals
CC   or transformed microorganisms. This sequence encodes the ORF2, ORF3 and
CC   the soxXYZA proteins from the P. pantotropus sox region.
XX
SQ   Sequence 3621 BP; 650 A; 1153 C; 1186 G; 632 T; 0 other;

  Query Match             1.8%;  Score 38.6;  DB 22;  Length 3621;
  Best Local Similarity   48.8%;  Pred. No. 2.3;
  Matches  104;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy    1847 TCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCG 1906
           ||||    |       |  | || || ||||  |     || ||| | ||   | |||| | |||
Db    3026 TCGAGGCCGCGATCCACGAAGACCATGGCGGGGTTGTCGAAATCGTCCCGCTCAAGATCC 2967

Qy    1907 CTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGA 1966
           |         |     |  |  | | ||| |        || |        |||   |||
Db    2966 CGCGTGCTGTCGTCGCGAAAGTGCCAGCCGGAATAGATCGTGTCGAACGTATCGGCCAGA 2907

Qy    1967 ACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCG 2026
              | |||  | | |   ||||||||  |   ||||   | || ||    |||  |   ||   |
Db    2906 AAGGCGGGCGGTGCGGTCTTGGTCACGATCTCGACCGGTCCGCTGTCGGTTTCGATGACC 2847

Qy    2027 AGTTCGAACCCGATCTGCAGCGCACAATCAGCG 2059
           ||  ||   | ||| | |    |||   |  | |||
Db    2846 AGCCCGTCCTCGACCGGGTCCGCGAAGGCCGCG 2814

RESULT 7
AAH48798/c
ID   AAH48798 standard; DNA; 4848 BP.
XX
AC   AAH48798;
XX
DT   05-NOV-2001  (first entry)
XX
DE   P. pantotrophus GB17 DNA encoding sox-associated proteins.
XX
KW   Sulphur oxidation; sox; ORF1; ORF2; ORF3; soxX; soxY; soxZ; soxA;
KW   reduced sulphur; thiosulphate; transgenic plant; transgenic animal; ds.
XX
OS   Paracoccus pantotrophus.
XX
PN   DE10006201-A1.
XX
PD   16-AUG-2001.
XX
PF   11-FEB-2000; 2000DE-1006201.
XX
PR   11-FEB-2000; 2000DE-1006201.
XX
PA   (FRIE/) FRIEDRICH C.
XX
PI   Friedrich C;
XX
DR   WPI; 2001-497622/55.
XX
PT   New genes involved in oxidation of reduced sulphur compounds,
PT   particularly conversion of thiosulphate to sulphate, and related
```

```
PT   proteins, from Paracoccus pantotrophus  -
XX
PS   Claim 1; Fig 1; 22pp; German.
XX
CC   This invention describes novel genes (I) of the sox (sulphur oxidation)
CC   region designated ORF1, ORF2, ORF3 (Open Reading Frame), and soxXYZA
CC   from Paracoccus pantotrophus GB17. The proteins (II) encoded by (I) are
CC   used for oxidation of reduced sulphur compounds in cellular or cell-free
CC   systems, particularly of thiosulphate to sulphate. (I) is also used for
CC   expression of the proteins and to prepare transgenic plants or animals
CC   or transformed microorganisms. This sequence encodes the P. pantotropus
CC   sox region associated proteins described in the invention.
XX
SQ   Sequence 4848 BP; 885 A; 1584 C; 1552 G; 827 T; 0 other;

  Query Match             1.8%;  Score 38.6;  DB 22;  Length 4848;
  Best Local Similarity   48.8%;  Pred. No. 2.7;
  Matches  104;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy   1847 TCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCG 1906
          ||||    |        |   | | || || ||||   |     || ||| | | ||    | |||| | |||
Db   4253 TCGAGGCCGCGATCCACGAAGACCATGGCGGGGTTGTCGAAATCGTCCCGCTCAAGATCC 4194

Qy   1907 CTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGA 1966
          |         |  |  |  | ||| |      || |         |||  |||
Db   4193 CGCGTGCTGTCGTCGCGAAAGTGCCAGCCGGAATAGATCGTGTCGAACGTATCGGCCAGA 4134

Qy   1967 ACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCG 2026
          | ||| | | |  |||||||| |  ||||  | || ||   ||| |  ||    ||    |
Db   4133 AAGGCGGGCGGTGCGGTCTTGGTCACGATCTCGACCGGTCCGCTGTCGGTTTCGATGACC 4074

Qy   2027 AGTTCGAACCCGATCTGCAGCGCACAATCAGCG 2059
          ||  ||  | ||| | |    |||  |  | |||
Db   4073 AGCCCGTCCTCGACCGGGTCCGCGAAGGCCGCG 4041


RESULT 8
ABL21478
ID   ABL21478 standard; DNA; 10757 BP.
XX
AC   ABL21478;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 15907.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 15907; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 10757 BP; 3125 A; 2268 C; 2226 G; 3138 T; 0 other;

  Query Match             1.8%;  Score 37.8;  DB 23;  Length 10757;
  Best Local Similarity   54.1%;  Pred. No. 7;
  Matches  100;  Conservative    0;  Mismatches   82;  Indels    3;  Gaps    1;

Qy   1283 AATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCG 1342
          | | | ||     | | |  || ||| |            ||  | |||| || ||||| | ||| |
Db   6709 ATTTGTTTTCGCAGATGAACCAAGTCCTGGATGCCATGTTCGAGCGAAAGGTTCAGCCGG 6768

Qy   1343 CCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGA 1402
               |  |||  | |||| | ||  |    | ||| ||  |     | |||     |
Db   6769 GT---GACTTTATCATCCGCCAGGGTGACGACGGCGATAACTTTTATGTTATTGAATCGT 6825

Qy   1403 ATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACA 1462
          | ||||   | || |||||  ||     || | | ||| |||||| ||||   | || | ||
Db   6826 AAGTAAAACAAAACAAATTTCAATCAAACTCCATGAATTATGCACACTTTCTGTTTGCCA 6885

Qy   1463 AGAGT 1467
          | |||
Db   6886 AAAGT 6890


RESULT 9
ABL18652/c
ID   ABL18652 standard; DNA; 9845 BP.
XX
AC   ABL18652;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 7429.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 7429; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
```

```
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 9845 BP; 2751 A; 2130 C; 2204 G; 2760 T; 0 other;

  Query Match             1.7%;  Score 37.4;  DB 23;  Length 9845;
  Best Local Similarity   49.7%;  Pred. No. 8.8;
  Matches   95;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy   1294 GACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGAT 1353
          |||||  | |  || |  |||   | | | | | | ||||  | | |  |  |    |||
Db   8624 GACAATAATACGGAGTGGCTCATTTGGATGGTCGGTAATATTCCGGGCTGCGATGTCGCC 8565

Qy   1354 ATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGAC 1413
          || |  |    || | |  |   |||||| |  |     ||   |   | |   ||| |
Db   8564 ATGGGTCAGACGCTTGTGGCCTACGACAACCGGCGGACCATCCACGGCAGTAACATCCAT 8505

Qy   1414 AAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTT 1473
             |   ||      | |   ||| |||||   ||  |||  |||||| ||| | | |||
Db   8504 CGAATCGTCTTCCTGGCCTTCAAGCAGTATCTGGAGCTCGATTTCGACGAGACTTTCGTT 8445

Qy   1474 CCTTATAGCGA 1484
          ||  |  ||||
Db   8444 CCCGAGGGCGA 8434


RESULT 10
AAC68978
ID   AAC68978 standard; DNA; 786 BP.
XX
AC   AAC68978;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-133 coding sequence.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria; ss.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
DR   P-PSDB; AAB37610.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 5; Fig 15; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The non-natural variants of the
CC   present invention are useful for immunising a mammal against malaria, and
CC   can be used to treat malaria. The present sequence is the MSP-133
CC   coding sequence.
XX
SQ   Sequence 786 BP; 245 A; 164 C; 159 G; 218 T; 0 other;

  Query Match             1.7%;  Score 37;  DB 22;  Length 786;
  Best Local Similarity   52.2%;  Pred. No. 3.1;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy   1377 CGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAA 1436
          ||  ||  ||   || || || ||      | || |||||| |   ||   | | |
Db    135 CGTTAACGTCAAGGACATTTTGAACTCCAGATTCAACAAGAGAGAAAACTTCAAGAACGT 194

Qy   1437 GAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGC 1496
             | |  | |||||  |     |||||  | ||  | ||| || | ||    |  | |
Db    195 TCTGGAGTCTGACTTGATTCCATACAAGGATTTGACTTCTTCTAACTACGTTGTTAAGGA 254

Qy   1497 CATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAG 1533
          |     ||||| | || ||||||| |||   | ||||
Db    255 CCCATACAAGTTCCTGAACAAGGAGAAGAGAGACAAG 291


RESULT 11
AAC68976
ID   AAC68976 standard; DNA; 1077 BP.
XX
AC   AAC68976;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Merozoite surface protein-142 coding sequence.
XX
KW   Merozoite surface protein; protazoacide; vaccine; malaria; ss.
XX
OS   Plasmodium falciparum.
XX
PN   WO200063245-A2.
XX
PD   26-OCT-2000.
XX
PF   20-APR-2000; 2000WO-GB01558.
XX
PR   20-APR-1999;   99GB-0009072.
PR   13-MAY-1999;   99US-0311817.
PR   25-MAY-1999;   99CA-2271451.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Holder A,  Birdsall B,  Feeney J,  Morgan W,  Syed S,  Uthaipibull C;
XX
DR   WPI; 2001-015762/02.
XX
PT   Novel variants of the C-terminal fragment of Plasmodium merozoite
PT   surface protein-1, useful as vaccines for treating or preventing
PT   malaria -
XX
PS   Example 5; Fig 15; 126pp; English.
XX
CC   The present invention relates to non-natural variants of a C-terminal
CC   fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC   non-natural variants have reduced affinity for at least 1 antibody
CC   capable of blocking a second antibody that inhibits the proteolytic
CC   cleavage of Plasmodium MSP-1_4_2, and has the same affinity for at least
CC   one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC   MSP-1_4_2, compared to natural MSP-1_1_9. The non-natural variants of the
CC   present invention are useful for immunising a mammal against malaria, and
CC   can be used to treat malaria. The present sequence is the MSP-142
CC   coding sequence.
XX
```

```
SQ   Sequence 1077 BP; 338 A; 228 C; 225 G; 286 T; 0 other;

  Query Match             1.7%;  Score 37;  DB 22;  Length 1077;
  Best Local Similarity   52.2%;  Pred. No. 3.6;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy   1377 CGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAA 1436
           || ||  ||   || || || ||        | || |||||| |    ||    | | |
Db    135 CGTTAACGTCAAGGACATTTTGAACTCCAGATTCAACAAGAGAGAAAACTTCAAGAACGT 194

Qy   1437 GAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGC 1496
               | |  | |||||  |       |||||  | ||   | |||  || | ||     | |
Db    195 TCTGGAGTCTGACTTGATTCCATACAAGGATTTGACTTCTTCTAACTACGTTGTTAAGGA 254

Qy   1497 CATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAG 1533
           |      ||||| | || ||||||| |||    | ||||
Db    255 CCCATACAAGTTCCTGAACAAGGAGAAGAGAGACAAG 291


RESULT 12
AAH01671/c
ID   AAH01671 standard; DNA; 810 BP.
XX
AC   AAH01671;
XX
DT   24-JUL-2001  (first entry)
XX
DE   Fonsecaea pedrosoi nucleotide sequence SEQ ID NO:1664.
XX
KW   Species specific; genus specific; family specific; probe; detection;
KW   identification; algal; archaeal; bacterial; fungal; parasitical;
KW   microorganism; diagnosis; translation elongation factor Tu; toxin;
KW   translation elongation factor G; RecA recombinase; resistance;
KW   catalytic subunit of proton-translocating ATPase; antimicrobial;
KW   vaccine; primer; ds.
XX
OS   Fonsecaea pedrosoi.
XX
PN   WO200123604-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-CA01150.
XX
PR   28-SEP-1999;   99CA-2283458.
PR   19-MAY-2000; 2000CA-2307010.
XX
PA   (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
XX
PI   Bergeron MG,  Boissinot M,  Huletsky A,  Menard C,  Ouellette M;
PI   Picard FJ,  Roy PH;
XX
DR   WPI; 2001-245006/25.
XX
PT   Nucleic acid sequences are used to generate universal probes and
PT   primers which can be used to identify and detect the presence of algal,
PT   archaeal, bacterial, fungal and parasitical species in a test sample -
XX
PS   Claim 24; Page 1264-1265; 1580pp; English.
XX
CC   The present invention describes a method for generating a repertory of
CC   nucleic acids of tuf, fus, atpD and/or recA genes from which probes
CC   and/or primers are derived. The method comprises amplifying the nucleic
CC   acids of determined algal, archaeal, bacterial, fungal and parasitical
CC   species with a combination of defined primer pairs. The method can be
CC   used for producing probes and/or primers for detecting one or more
CC   related microorganisms e.g. algae, archaea, bacteria, fungi and
CC   parasites, for universal detection and for specific and ubiquitous
CC   detection and identification of an algal, archaeal, bacterial, fungal
CC   and parasitical species, genus, family and group. A nucleic acid (I)
CC   obtained using the method of the invention can be used for the universal
CC   detection of any bacterium, fungus or parasite in a sample and for the
CC   detection of at least one antimicrobial agent resistance gene or at
CC   least one toxin gene. hexA nucleic acids are used for the specific and
CC   ubiquitous detection and for identification of Streptococcus pneumoniae.
CC   (I) can be used to design a therapeutic agent which is effective against
CC   microorganisms. Microbial species or genus or family or phylum or group
CC   which can be detected include Abiotrophia adiacens, Bordetella sp.,
CC   Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
CC   Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,
CC   Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
CC   provides faster results than substrate specificity tests as results can
CC   be determined in an hour and improved accuracy is also achieved.
CC   AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
CC   which are given in the exemplification of the present invention.
XX
SQ   Sequence 810 BP; 167 A; 254 C; 228 G; 161 T; 0 other;

  Query Match             1.7%;  Score 36.8;  DB 22;  Length 810;
  Best Local Similarity   46.2%;  Pred. No. 3.6;
  Matches  122;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy    192 TGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGAT 251
            || ||  || |   |||| || ||   | | ||| ||    |||  |||  ||  | | |
Db    305 TGGCGATGTCACGCTCGGGGGTGGGGATCCAGGTGTCGACGTGCTCGAGGAGCTCGTCAA 246

Qy    252 CTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGA 311
             ||  | || |  | |   |||   ||  ||  || |  || ||    |      || ||
Db    245 TCTTCTCGACACCAATGTCGGGCTCGCGGCCCTCAATGGCGCAAAGGGCGGAACCCATGA 186

Qy    312 CTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCT 371
            | ||  | | |  |      | |       |        || ||| |||| ||  ||   |||
Db    185 CGATGGGAGTGTCGTCACCCTCGAAGCCGTAGCTGGAGAGAAGCTCACGCATCTCCATCT 126

Qy    372 TTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACAGCTCAA 431
               || |    | | |  | |     || | || |   |  |||   || || | | |   |||
Db    125 CGACGAGCTCCAACATCTCCTTGTCCTCAATGGCATCGACCTTGTTGACGAAGACGACAA 66

Qy    432 GGGTATCACTGACGAGATGGAGCG 455
            |   |        |||  ||| | |
Db     65 TGCGCTTGACACCGACCTGGCGGG 42


RESULT 13
AAF86431/c
ID   AAF86431 standard; DNA; 349980 BP.
XX
AC   AAF86431;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Pyrococcus abyssi genomic fragment #1.
XX
KW   Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX
OS   Pyrococcus abyssi.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    300001..349980
FT                   /*tag=  a
FT                   /note= "This sequence overlaps with the 5' end of
FT                   AAH41223"
XX
PN   FR2792651-A1.
XX
PD   27-OCT-2000.
XX
PF   21-APR-1999;   99FR-0005034.
XX
PR   21-APR-1999;   99FR-0005034.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
PA   (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI   Forterre P,  Thierry JC,  Prieur D,  Dietrich J,  Lecompte O;
PI   Querellou J,  Weissenbach J,  Saurin W,  Heilig R;
```

```
XX
DR   WPI; 2001-126236/14.
XX
PT   New nucleotide sequences isolated from Pyrococcus abyssi encode
PT   proteins useful in industry -
XX
PS   Claim 1; Page 183-279; 1657pp; French.
XX
CC   The present invention relates to the genomic sequence of Pyrococcus
CC   abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
CC   hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
CC   vents. The present sequence is a fragment of the genomic sequence of P.
CC   abyssi. The 3' end of this sequence overlaps with the 5' end of AAH41223.
CC   The proteins of the present invention have various potential industrial
CC   uses, since the proteins are stable at very high temperatures, some up to
CC   110 degrees centigrade.
CC   Note: This patent is in the same patent family as WO200065062, which
CC   contains additional sequences as shown in AAB99132-AAB99143,
CC   AAH75903-AAH75920 and AAG66436.
XX
SQ   Sequence 349980 BP; 95167 A; 78444 C; 78870 G; 97499 T; 0 other;

  Query Match             1.7%;  Score 36.8;  DB 22;  Length 349980;
  Best Local Similarity   49.5%;  Pred. No. 84;
  Matches   95;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy    1377 CGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAA 1436
            |||  ||||     |||  |        || | ||||     ||        ||    |   | |
Db  168164 CGATGAGCTTGAGGATGAAGAGCTTAAGGAAATCCTGGAGGTGGCTGAAAAGGTTGCAAG 168105

Qy    1437 GAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGC 1496
            | |   | | |||| |||    ||     ||||  ||     |     |||  || | | |||  ||
Db  168104 GGAAAAAGGAGACGTCGAGCTCTTAAAGATGGTAACTTACTATTACGCCGAATTCTTTGG 168045

Qy    1497 CATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGT 1556
             ||     |||    | |||     ||| |||| |||||||  |    |  |||||           |
Db  168044 GATAGATAAGCTAAAGGAGTTTGAAGAGTTCGCCAAGGGCAAGGGAAAAGAGGGACTTTT 167985

Qy    1557 AGAGCTTTCCAA 1568
              ||||   || |
Db  167984 CCAGCTCGCCGA 167973


RESULT 14
AAC66137
ID   AAC66137 standard; DNA; 1090 BP.
XX
AC   AAC66137;
XX
DT   13-FEB-2001  (first entry)
XX
DE   Human kidney disease associated gene SEQ ID 7.
XX
KW   Nephrotropic; cytostatic; human; kidney disease associated; hypertension;
KW   Bartter's syndrome; Gitelman syndrome; nephrolithiasis; renal failure;
KW   renal amyloidosis; primary aldosteronism; Addison's disease; cancer;
KW   glomerulonephritis; dysplastic malformation; medullary cystic disease;
KW   medullary sponge kidney; tubular dysplasia; Alport's syndrome; ds.
XX
OS   Homo sapiens.
XX
PN   WO200061622-A2.
XX
PD   19-OCT-2000.
XX
PF   28-MAR-2000; 2000WO-US08260.
XX
PR   09-APR-1999;  99US-0289349.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Walker MG,  Volkmuth W,  Klingler TM,  Azimzai Y,  Yue H;
XX
```

```
DR   WPI; 2000-665116/64.
DR   P-PSDB; AAY85680.
XX
PT   Novel kidney disease associated gene useful for diagnosing treating and
PT   preventing renal disorders, is uromodulin, NKCC2, NCCT, aldolase B,
PT   ROMK1, ATP1G1, PDZK1, NPT1, calbindin, kininogen or ClC-Kb  -
XX
PS   Claim 2; Page 33-34; 36pp; English.
XX
CC   Polynucleotides AAC66131-C66139 represent human kidney disease
CC   associated genes. Proteins AAY85678-Y85680 represent human kidney disease
CC   associated proteins. The polynucleotides have nephrotropic and cytostatic
CC   activity. The polynucleotides, encoded proteins and pharmaceutical
CC   compositions containing them are useful for diagnosing, prognosing,
CC   treating and preventing renal disorders such as Bartter's syndrome,
CC   Gitelman syndrome, autosomal dominant polycystic kidney disease and
CC   nephrolithiasis. The genes and proteins are also useful for identifying
CC   biomolecules that are associated with a specific disease, regulatory
CC   pathway, subcellular compartment, cell type, tissue type or species,
CC   which is useful in diagnosis, prognosis, treatment and evaluation of
CC   therapies for renal diseases for e.g. renal amyloidosis, hypertension,
CC   primary aldosteronism, Addison's disease, renal failure,
CC   glomerulonephritis, chronic glomerulonephritis, tubulointerstitial
CC   nephritis, cystic disorders and dysplastic malformations, inherited
CC   polycystic renal disease (PRD), medullary cystic disease, medullary
CC   sponge kidney and tubular dysplasia, Alport's syndrome, non-renal
CC   cancers, kidney adenocarcinoma, metastatic renal carcinoma, multiple
CC   myeloma and nephrotoxic disorders.
XX
SQ   Sequence 1090 BP; 215 A; 364 C; 303 G; 208 T; 0 other;

  Query Match             1.7%;  Score 36.4;  DB 21;  Length 1090;
  Best Local Similarity   49.5%;  Pred. No. 5.5;
  Matches   94;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy       1674 CCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCAT 1733
               ||  |  ||||   |   |  |  |  |       |  |    ||  ||     ||     | || | |
Db        136 CCTGGACCGTGTCTTCACCACCTACAAGCTCATGCACACGCACCAGACAGTGGACTTCGT 195

Qy       1734 CAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGT 1793
               || |  |   |   |   |  |||      || |  ||   |  |  |  |   |  | ||||| || ||   |
Db        196 CAGGAGCAAGCATGCCCAGTTTGGGGGCTTCTCCTACAAGAAAATGACAGTCATGGAGGC 255

Qy       1794 ATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCT 1853
                 ||||    || ||||       |      |||||||   | |       | |     || ||| | ||
Db        256 CGTGGACCTGCTGGATGGGCTGGTGGATGAGTCGGACCCGGACGTAGATTTCCCCAACTC 315

Qy       1854 CTTCCGCACC 1863
               |||||    ||
Db        316 CTTCCATGCC 325


RESULT 15
AAI58976
ID   AAI58976 standard; cDNA; 1292 BP.
XX
AC   AAI58976;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 1179.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
```

```
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM39820.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 1179; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 1292 BP; 254 A; 422 C; 368 G; 248 T; 0 other;

  Query Match             1.7%;  Score 36.4;  DB 22;  Length 1292;
  Best Local Similarity   49.5%;  Pred. No. 6;
  Matches   94;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy   1674 CCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCAT 1733
          ||  |  ||||   |  | | | |      | |   || ||    ||       | || | |
Db    339 CCTGGACCGTGTCTTCACCACCTACAAGCTCATGCACACGCACCAGACAGTGGACTTCGT 398

Qy   1734 CAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGT 1793
          || | |  |  |  | |||   || | ||  | | || |  | | |||||| || ||  | 
Db    399 CAGGAGCAAGCATGCCCAGTTTGGGGGCTTCTCCTACAAGAAAATGACAGTCATGGAGGC 458

Qy   1794 ATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCT 1853
            ||||   || ||||     |   |||||||  | |    | |   ||   ||| | ||  
Db    459 CGTGGACCTGCTGGATGGGCTGGTGGATGAGTCGGACCCGGACGTAGATTTCCCCAACTC 518

Qy   1854 CTTCCGCACC 1863
          |||||   ||
Db    519 CTTCCATGCC 528

Search completed: May 23, 2003, 04:52:58
Job time : 958 secs
```

```
                    GenCore version 5.1.4_p5_4578
           Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         May 23, 2003, 02:43:56 ; Search time 62 Seconds
                                          (without alignments)
                                          10580.348 Million cell updates/sec

Title:          US-10-008-355-1
Perfect score:  2139
Sequence:       1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters:      882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                 1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                 2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                 3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                 4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                 5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                 6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1717.4	80.3	1974	4	US-09-221-017B-726	Sequence 726, App
2	207.4	9.7	1317	4	US-09-221-017B-382	Sequence 382, App
3	164.4	7.7	561	4	US-09-221-017B-12	Sequence 12, Appl
4	153.8	7.2	2384	4	US-09-221-017B-1045	Sequence 1045, Ap
5	36.4	1.7	1090	4	US-09-289-349-7	Sequence 7, Appli
6	36.2	1.7	2280	1	US-08-415-818-5	Sequence 5, Appli
7	36.2	1.7	2280	2	US-08-894-236-5	Sequence 5, Appli
8	36.2	1.7	2280	5	PCT-US96-01444-5	Sequence 5, Appli
c 9	36	1.7	7218	1	US-08-232-463-14	Sequence 14, Appl
10	35.2	1.6	449	2	US-08-825-782-2	Sequence 2, Appli
c 11	34	1.6	5837	1	US-07-686-340-1	Sequence 1, Appli
c 12	34	1.6	5837	1	US-08-004-139B-1	Sequence 1, Appli
c 13	34	1.6	5837	1	US-08-117-491-1	Sequence 1, Appli
c 14	34	1.6	5837	1	US-08-271-364A-1	Sequence 1, Appli
c 15	34	1.6	5837	2	US-08-811-492-1	Sequence 1, Appli
c 16	34	1.6	5837	2	US-08-222-715B-1	Sequence 1, Appli
c 17	34	1.6	5837	5	PCT-US96-10545A-1	Sequence 1, Appli
18	33.6	1.6	23673	4	US-09-773-816-1	Sequence 1, Appli
19	33.6	1.6	50937	4	US-09-428-517-1	Sequence 1, Appli
20	33.2	1.6	939	4	US-09-105-390-45	Sequence 45, Appl
21	33.2	1.6	1020	4	US-09-105-390-61	Sequence 61, Appl
22	33.2	1.6	2612	4	US-09-105-390-7	Sequence 7, Appli
23	32.4	1.5	710	4	US-08-998-416-603	Sequence 603, App
24	32.2	1.5	3468	1	US-07-951-715A-2	Sequence 2, Appli
25	32.2	1.5	3468	2	US-08-459-448A-2	Sequence 2, Appli
26	32.2	1.5	3468	3	US-08-459-595A-2	Sequence 2, Appli
27	32.2	1.5	3468	3	US-08-459-504B-2	Sequence 2, Appli
28	32.2	1.5	3468	3	US-08-459-444-2	Sequence 2, Appli
29	32.2	1.5	3468	3	US-09-053-549-3	Sequence 3, Appli
30	32.2	1.5	3468	4	US-09-547-422-2	Sequence 2, Appli
31	32	1.5	289	4	US-09-007-005-17	Sequence 17, Appl
32	32	1.5	289	4	US-09-244-796-17	Sequence 17, Appl
c 33	32	1.5	823	4	US-08-998-416-551	Sequence 551, App
34	32	1.5	1947	1	US-07-951-715A-3	Sequence 3, Appli
35	32	1.5	1947	2	US-08-459-448A-3	Sequence 3, Appli
36	32	1.5	1947	3	US-08-459-595A-3	Sequence 3, Appli
37	32	1.5	1947	3	US-08-459-504B-3	Sequence 3, Appli
38	32	1.5	1947	3	US-08-459-444-3	Sequence 3, Appli
39	32	1.5	1947	4	US-09-547-422-3	Sequence 3, Appli
40.	32	1.5	2845	1	US-08-289-653-1	Sequence 1, Appli
41	32	1.5	3468	1	US-07-951-715A-4	Sequence 4, Appli
42	32	1.5	3468	1	US-07-951-715A-8	Sequence 8, Appli
43	32	1.5	3468	2	US-08-459-448A-4	Sequence 4, Appli
44	32	1.5	3468	2	US-08-459-448A-8	Sequence 8, Appli
45	32	1.5	3468	3	US-08-459-595A-4	Sequence 4, Appli

```
                                ALIGNMENTS

RESULT 1
US-09-221-017B-726
; Sequence 726, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  726:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1974 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
```

```
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:   1...1974
US-09-221-017B-726

  Query Match             80.3%;  Score 1717.4;  DB 4;  Length 1974;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1730;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    1;

Qy     409 GACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCT 468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCT 60

Qy     469 CAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATC 528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATC 120

Qy     529 GTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGAC 588
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGAC 180

Qy     589 GTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAAC 648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAAC 240

Qy     649 TGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGAC 708
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGAC 300

Qy     709 AACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCC 768
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCC 360

Qy     769 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 828
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 420

Qy     829 ACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCT 888
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCT 480

Qy     889 CGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAG 948
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAG 540

Qy     949 GCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCG 1008
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCG 600

Qy    1009 ATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAA 1068
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAA 660

Qy    1069 AGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTG 1128
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTG 720

Qy    1129 TCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTG 1188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTG 780

Qy    1189 AGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTG 1248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 AGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTG 840

Qy    1249 GCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGAC 1308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGAC 900

Qy    1309 TACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGT 1368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGT 960

Qy    1369 ATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC 1428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC 1020

Qy    1429 GACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAG 1488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAG 1080

Qy    1489 TTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGAT 1548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGAT 1140

Qy    1549 CCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCG 1608
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCG 1200

Qy    1609 ATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAG 1668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAG 1260

Qy    1669 ATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGC 1728
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGC 1320

Qy    1729 TCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAG 1788
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAG 1380

Qy    1789 GGCGTATTGGAGAAGCAGGAT--CCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCC 1846
           |||||||||||||||||||||   | ||||||||||||||||||||||||||||||||||
Db    1381 GGCGTATTGGAGAAGCAGGATTCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCC 1440

Qy    1847 TCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCG 1906
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCG 1500

Qy    1907 CTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGA 1966
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGA 1560

Qy    1967 ACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCG 2026
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 ACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCG 1620

Qy    2027 AGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGA 2086
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 AGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGA 1680

Qy    2087 TTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 2139
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 TTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 1733


RESULT 2
US-09-221-017B-382
; Sequence 382, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  382:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1317 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...1317
US-09-221-017B-382

  Query Match             9.7%;  Score 207.4;  DB 4;  Length 1317;
  Best Local Similarity   53.1%;  Pred. No. 5e-56;
  Matches  498;  Conservative    0;  Mismatches  421;  Indels   18;  Gaps    2;

Qy    111 GAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTT 170

Db      6 GAAGTATGCGCAGATGAAAGAGCGCGGCCTTAAGATGAAGGAATACGACCTTTATAATCC 65

Qy    171 CGACAAGCCGTCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCAC 230

Db     66 CAACGGCACATCGCTGAAAGATGCCGTAGTGCTATTCGACGGAGGATGTACGGGCGAGGT 125

Qy    231 AGTGTCCGATCAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAG 290

Db    126 CGTTTCCGATCGCGGACTGGTACTGACCAATCACCACTGCGGATACGATATGATCCAGGC 185

Qy    291 CCAAAGCACGGTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGA 350

Db    186 TCACAGCACGCTCGAGCATAACTATCTCGAAAATGGATTTTGGGCGATGAGAGAAGCGGA 245

Qy    351 GGAGCTTCCGATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGA 410

Db    246 TGAATTACCGAACAAGGATATTTCCGTGGTATTCATCGACAAGATCGAAGATGTCACAGA 305

Qy    411 CAAGGTAGAAGGACAGCTCAAGGGTATCACTGA------------CGAGATGGAGCGTCT 458

Db    306 CTACGTCAAGAAAGATCTCAAAGCCATCAAAGATCCCAACAGCATGGACTACCTCTCTCC 365

Qy    459 GCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCA 518

Db    366 GAAGTACCTGCAAAAGTTGGCTGACAAGAAGGCCGGCAAGAACTTTTCTGCCAAGAATCC 425

Qy    519 ACTCTGCATCGTAGAG------CCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTA 572

Db    426 GGGGCTTTCCGTAGAGATCAAAGCCTTCTATGGGGCAATCTCTACCTGATGTTTACCAA 485

Qy    573 CGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGG 632

Db    486 AAAGACTTATACGGATGTTCGACTGGTGGGAGCACCTCCCACCAGCATTGGCAAATTCGG 545

Qy    633 AGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGT 692

Db    546 TGCCGATACGGACAACTGGATCTGGCCTCGTCATACTGGCGACTTCTCCATCTTCCGTAT 605

Qy    693 GTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCC 752

Db    606 CTATGCGGACAAGAATGGCAATCCCGCACCATACTCTGAAGATAATGTTCCGCTCAAGCC 665

Qy    753 CGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCAT 812

Db    666 GAAGCGTTTCTTCAATATCTCCCTTGGTGGAGTACAAGAGAACGACTACGCCATGATAAT 725

Qy    813 CGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGA 872

Db    726 GGGTTTCCCCGGTACTACGCACCGCTATTTCACGGCTTCCGAAGTAGACGAATGGAAAAG 785

Qy    873 AAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGC 932

Db    786 CATCGACAACGATATTCGCATCCGCATGCGTGATATTCGTCAGGGTGTCATGCTCAGGGA 845

Qy    933 CATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAA 992

Db    846 AATGCTGGCCGATCCTCAGATCAAAATCATGTATTCAGCTAAATATGCCGCTTCGCAGAA 905

Qy    993 CTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTC 1029

Db    906 TGCTTACAAACGTGCTATAGGTGCCAACTGGGCGATC 942


RESULT 3
US-09-221-017B-12
; Sequence 12, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
```

```
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...561
US-09-221-017B-12

  Query Match             7.7%;  Score 164.4;  DB 4;  Length 561;
  Best Local Similarity   95.7%;  Pred. No. 1.7e-42;
  Matches  180;  Conservative    0;  Mismatches    6;  Indels    2;  Gaps    1;

Qy      1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    366 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTAAGGG 425

Qy     61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 485

Qy    121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
          ||||||||||||||||||||  |  ||  |||||||||||||||||||||||||||||||
Db    486 CGAATGCGTGAGCTCGGCTTTA--CCTCCCGTGGATTCGCTCTACAGTTTCGACAAGCCG 543

Qy    181 TCCATTGC 188
          ||||| ||
Db    544 TCCATGGC 551

RESULT 4
US-09-221-017B-1045
; Sequence 1045, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  1045:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2384 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...2384
US-09-221-017B-1045

  Query Match             7.2%;  Score 153.8;  DB 4;  Length 2384;
  Best Local Similarity   55.0%;  Pred. No. 1e-38;
  Matches  324;  Conservative    0;  Mismatches  262;  Indels    3;  Gaps    1;

Qy   1544 AAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTCGCGCTATTCAGGCCG 1603
          ||||||    |||    |  ||     ||  |  |    |||     |  |   ||
Db     31 AAGATCCTATGGTACTCTTCGCCTCTTCGGTTTTCGACGAATACCGCAAACTCTACAACG 90

Qy   1604 ATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGC 1663
          |         || ||| | || | |||    | ||    | ||     | | |||| ||| |
Db     91 AACTTCGTCCCTATGACGATCCTATTCTAAGGGCACAACGCACCTACATTGCAGGTCTCT 150

Qy   1664 GTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCT 1723
            || ||| |    ||   |   | ||           ||||| ||| | ||  | ||| | | ||
Db    151 TGGAAATGGATGGCGATCAGGATCAATTCCCGGATGCTAACCTGACACTTCGTTTCACCT 210

Qy   1724 ACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAG 1783
          | ||     | ||||| |||  |||  |    |||   ||     |||     |||| || |||
Db    211 ATGGTCAAGTGAAGGGCTATTCACCCCGTGACAATGTTTACTACGGACATCAAACCACAT 270

Qy   1784 GCAAGGGCGTATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATA 1843
             | || ||  |||| ||  | |||||  | |       || ||||  ||   |     ||
Db    271 TGGATGGTGTGATGGAAAAAGAAGATCCCGATAATTGGGAATTTGTAGTCGATCCCAAGC 330

Qy   1844 TCCTCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGA---GAACGGTCAGCTCC 1900
          |    | | | | |       ||| |||  || || |||||||     | |||||   | |
Db    331 TGAAAGCCGTATACGAGCGTAAAGACTTCGGGCGTTATGCCGATCGCAGCGGTCGCATGC 390

Qy   1901 ATATCGCTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCG 1960
            | | || ||      | | ||    | |   || ||||| ||||| || || || ||   |
Db    391 CTGTAGCCTTTTGCGCCACCACACATACAACCGGCGGCAACTCAGGCAGTCCGGTCATGA 450

Qy   1961 ATAAGAACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTG 2020
          ||    ||||||    |||||||||||     |||||| | ||||||||  |    |  |||| |
Db    451 ATGCCAACGGCGAACTGATCGGTCTCAACTTCGATCGTAACTGGGAGGGAGTCGGTGGCG 510

Qy   2021 ACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCT 2080
```

```
              ||||| |||        ||||     ||||||| ||||  ||||| || |||||||| ||
Db     511 ACATCCAGTATCTGGCCGACTACCAGCGCAGCATCATTGTGGATATTCGCTACGTCCTGC 570

Qy    2081 TCATGATTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGCTGAA 2129
              |  |||| ||||||   ||      ||||  || ||   |   | ||   ||||
Db     571 TTGTGATAGACAAAGTAGGCGGTTGCCAACGCCTGTTGGATGAAATGAA 619


RESULT 5
US-09-289-349-7
; Sequence 7, Application US/09289349
; Patent No. 6277574
; GENERAL INFORMATION:
;  APPLICANT: Walker, Michael, G.
;  APPLICANT: Volkmuth, Wayne
;  APPLICANT: Klinger, Tod, M.
;  APPLICANT: Azimzai, Yalda
;  APPLICANT: Yue, Henry
;  TITLE OF INVENTION: GENES ASSOCIATED WITH DISEASES OF THE KIDNEY
;  FILE REFERENCE: PB-0010 US
;  CURRENT APPLICATION NUMBER: US/09/289,349
;  CURRENT FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PERL Program
; SEQ ID NO 7
;   LENGTH: 1090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 2673009CB1
US-09-289-349-7

  Query Match             1.7%;  Score 36.4;  DB 4;  Length 1090;
  Best Local Similarity   49.5%;  Pred. No. 0.23;
  Matches   94;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy    1674 CCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCAT 1733
              || |  ||||   |  | | | |        | |   || ||    ||     | || | |
Db     136 CCTGGACCGTGTCTTCACCACCTACAAGCTCATGCACACGCACCAGACAGTGGACTTCGT 195

Qy    1734 CAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGT 1793
              || | |   |   |   | |||    || | ||   | | || |   | | ||||| || ||   |
Db     196 CAGGAGCAAGCATGCCCAGTTTGGGGGCTTCTCCTACAAGAAAATGACAGTCATGGAGGC 255

Qy    1794 ATTGGAGAAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCT 1853
               ||||   || ||||         |      |||||||   | |       | |     || ||| | ||
Db     256 CGTGGACCTGCTGGATGGGCTGGTGGATGAGTCGGACCCGGACGTAGATTTCCCCAACTC 315

Qy    1854 CTTCCGCACC 1863
              |||||    ||
Db     316 CTTCCATGCC 325


RESULT 6
US-08-415-818-5
; Sequence 5, Application US/08415818
; Patent No. 5621079
;  GENERAL INFORMATION:
;    APPLICANT:  Cascieri, Margaret A.
;    APPLICANT:  Linemeyer, David L.
;    APPLICANT:  MacNeil, Douglas J.
;    APPLICANT:  Shiao, Lin-Lin
;    APPLICANT:  Strader, Catherine D.
;    APPLICANT:  Tan, Carina P.
;    APPLICANT:  Weinberg, David H.
;    TITLE OF INVENTION:  NEUROPEPTIDE Y RECEPTOR
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mary A. Appollina
;      STREET:  P.O. Box 2000, 126 E. Lincoln Ave.
;      CITY:  Rahway
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/415,818
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/383,746
;      FILING DATE:  03-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Appollina, Mary A.
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19390
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  908-594-3462
;      TELEFAX:  908-594-4720
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  822..1937
US-08-415-818-5

  Query Match             1.7%;  Score 36.2;  DB 1;  Length 2280;
  Best Local Similarity   55.0%;  Pred. No. 0.44;
  Matches   71;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy     409 GACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCT 468
           |||||||  ||      |||| ||||   ||||  |       | ||  |  | ||  ||  |  ||||||
Db     501 GACAAGGAGGATCAGAAGCTTAAGGACATCATTTTGTACATAGTGAGTTTGAGGAAAGCT 560

Qy     469 CAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATC 528
             |||     |||   |          ||   |      ||| | ||   || |    | |||  ||    ||| |   |
Db     561 GAGGTTACATGGAACTCTCTCTCTCAAAAACAAAACAAAACAAAACAAAACCTTCTAC 620

Qy     529 GTAGAGCCT 537
              |  |   ||
Db     621 TAATATTCT 629


RESULT 7
US-08-894-236-5
; Sequence 5, Application US/08894236
; Patent No. 5939263
;  GENERAL INFORMATION:
;    APPLICANT:  Cascieri, Margaret A.
;    APPLICANT:  Linemeyer, David L.
;    APPLICANT:  MacNeil, Douglas J.
;    APPLICANT:  Shiao, Lin-Lin
;    APPLICANT:  Strader, Catherine D.
;    APPLICANT:  Tan, Carina P.
;    APPLICANT:  Weinberg, David H.
;    TITLE OF INVENTION:  NEUROPEPTIDE Y RECEPTOR
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mary A. Appollina
;      STREET:  P.O. Box 2000, 126 E. Lincoln Ave.
;      CITY:  Rahway
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/894,236
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/383,746
;      FILING DATE:  03-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/415,818
;      FILING DATE:  03-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Appollina, Mary A.
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19390Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  908-594-3462
;      TELEFAX:  908-594-4720
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  822..1937
US-08-894-236-5

  Query Match             1.7%;  Score 36.2;  DB 2;  Length 2280;
  Best Local Similarity   55.0%;  Pred. No. 0.44;
  Matches   71;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy    409 GACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCT 468
          |||||||  ||     |||| ||||  |||| |     | || | | | || || | ||||||
Db    501 GACAAGGAGGATCAGAAGCTTAAGGACATCATTTTGTACATAGTGAGTTTGAGGAAAGCT 560

Qy    469 CAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATC 528
           |||    |||  |     ||  |     ||| | ||  || |  | ||| ||  || |  |
Db    561 GAGGTTACATGGAACTCTCTCTCTCTCAAAAACAAAACAAAACAAAACAAAACCTTCTAC 620

Qy    529 GTAGAGCCT 537
            | |  ||
Db    621 TAATATTCT 629


RESULT 8
PCT-US96-01444-5
; Sequence 5, Application PC/TUS9601444
;  GENERAL INFORMATION:
;    APPLICANT:  Cascieri, Margaret A.
;    APPLICANT:  Linemeyer, David L.
;    APPLICANT:  MacNeil, Douglas J.
;    APPLICANT:  Shiao, Lin-Lin
;    APPLICANT:  Strader, Catherine D.
;    APPLICANT:  Tan, Carina P.
;    APPLICANT:  Weinberg, David H.
;    TITLE OF INVENTION:  NEUROPEPTIDE Y RECEPTOR
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mary A. Appollina
;      STREET:  P.O. Box 2000, 126 E. Lincoln Ave.
;      CITY:  Rahway
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/01444
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/383,746
;      FILING DATE:  03-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/415,818
;      FILING DATE:  03-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Appollina, Mary A.
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19390Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  908-594-3462
;      TELEFAX:  908-594-4720
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  822..1937
PCT-US96-01444-5

  Query Match             1.7%;  Score 36.2;  DB 5;  Length 2280;
  Best Local Similarity   55.0%;  Pred. No. 0.44;
  Matches   71;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy    409 GACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCT 468
          |||||||  ||     |||| ||||  |||| |     | || | | | || || | ||||||
Db    501 GACAAGGAGGATCAGAAGCTTAAGGACATCATTTTGTACATAGTGAGTTTGAGGAAAGCT 560

Qy    469 CAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATC 528
           |||    |||  |     ||  |     ||| | ||  || |  | ||| ||  || |  |
Db    561 GAGGTTACATGGAACTCTCTCTCTCTCAAAAACAAAACAAAACAAAACAAAACCTTCTAC 620

Qy    529 GTAGAGCCT 537
            | |  ||
Db    621 TAATATTCT 629


RESULT 9
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14

  Query Match             1.7%;  Score 36;  DB 1;  Length 7218;
  Best Local Similarity   3.2%;  Pred. No. 1.1;
  Matches   12;  Conservative  204;  Mismatches  164;  Indels    0;  Gaps    0;

Qy     851 GGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCA 910
           :::: : ::::::  : :  ::::: ::::: ::      : :  ::::   : :: :  :
Db    1423 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1364

Qy     911 AGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCA 970
           :: :::: :    ::::::::  : ::: : :::  :::  :    : :  ::: : :  :
Db    1363 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1304

Qy     971 GCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCG 1030
           : ::: : :    :::: :  ::  :  :::::::   ::  :: : :::  : ::    :
Db    1303 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1244

Qy    1031 CTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCC 1090
             :      :: : :: :::  : ::: : :  ::::::::: :    : :::  :::
Db    1243 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1184

Qy    1091 GTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATA 1150
           : ::::: :: ::::: :   :    : :: :: : :  :            ::::::   : :
Db    1183 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1124

Qy    1151 AGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTA 1210
           ::::::::  ::::  ::   :  :::: ::   :    ::: :::: :        :: : |
Db    1123 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATC 1064

Qy    1211 CCGAGGTGGTTCGTTTTGCA 1230
             | || |   |||   ||||
Db    1063 GCAAGCTCCCTCGACCTGCA 1044


RESULT 10
US-08-825-782-2
; Sequence 2, Application US/08825782
; Patent No. 5834239
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  COFACTOR A-LIKE PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
```

```
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/825,782
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0263 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  449 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  PROSTUT08
;      CLONE:  1654275
US-08-825-782-2

  Query Match             1.6%;  Score 35.2;  DB 2;  Length 449;
  Best Local Similarity   53.7%;  Pred. No. 0.32;
  Matches   73;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     422 GACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCC 481
           ||||| |||||    |  || | || | || ||||  ||  |||||   | | | | |
Db      89 GACAGATCAAGATCAAGACCGGCGTGGTGAAGCGGTTGGTCAAAGAAAAAGTGATGTATG 148

Qy     482 AAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCT 541
           || ||  ||| ||| || || | | | |       |  || |  |     | |||   |
Db     149 AAAAAGAGGCAAAACAACAAGAAGAAAAGATTGAAAAAATGAGAGCTGAAGACGGTGAAA 208

Qy     542 ATTCCAACAACGAATA 557
           ||| | |||    || |
Db     209 ATTACGACATTAAAAA 224


RESULT 11
US-07-686-340-1/c
; Sequence 1, Application US/07686340
; Patent No. 5322785
;  GENERAL INFORMATION:
;    APPLICANT:  Comb, Donald G.
;    APPLICANT:  Perler, Francine
;    APPLICANT:  Kucera, Rebecca
;    APPLICANT:  Jack, William E.
;    TITLE OF INVENTION:  Purified Thermostable DNA Polymerase
;    TITLE OF INVENTION:  Obtainable From Thermococcus Litoralis
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS
;      ADDRESSEE:  & CUSHMAN
;      STREET:  130 WATER STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/686,340
;       FILING DATE:  19910417
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/513,994
;       FILING DATE:  26-APR-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/626,057
;       FILING DATE:  11-DEC-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  RESNICK, DAVID S.
;       REGISTRATION NUMBER:  34235
;       REFERENCE/DOCKET NUMBER:  39296-CIP-II
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 523-3400
;       TELEFAX:  (617) 523-6440
;       TELEX:  200291 STRE UR
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5837 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-07-686-340-1

  Query Match             1.6%;  Score 34;  DB 1;  Length 5837;
  Best Local Similarity   57.5%;  Pred. No. 4.1;
  Matches   61;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy   1405 GTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAG 1464
            || |||| | |||  |         | ||||     || ||||||  ||| | |||| |||
Db   3686 GTCATCGTCCAGAGTTAGTGCTTCAACACCTTCGAGAATGCAGTATTCTTTTTCGCCAAT 3627

Qy   1465 AGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCA 1510
             ||| || |     ||| || ||| ||| |     | | ||| || |
Db   3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTCACAAATCTA 3581


RESULT 12
US-08-004-139B-1/c
; Sequence 1, Application US/08004139B
; Patent No. 5496714
;  GENERAL INFORMATION:
;     APPLICANT:  COMB, DONALD G.
;     APPLICANT:  PERLER, FRANCINE B.
;     APPLICANT:  XU, MING-QUN
;     APPLICANT:  HODGES, ROBERT A.
;     APPLICANT:  JACK, WILLIAM E.
;     TITLE OF INVENTION:  MODIFIED PROTEINS AND METHODS OF THEIR
;     TITLE OF INVENTION:  PRODUCTION
;     NUMBER OF SEQUENCES:  39
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;       ADDRESSEE:  CUSHMAN
;       STREET:  130 WATER STREET
;       CITY:  BOSTON
;       STATE:  MASSACHUSETTS
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/004,139B
;       FILING DATE:  09-DEC-1992
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  RESNICK, DAVID S.
;       REGISTRATION NUMBER:  34235
;       REFERENCE/DOCKET NUMBER:  42188
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 523-3400
;       TELEFAX:  (617) 523-6440
;       TELEX:  200291 STRE UR
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5837 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  not relevant
US-08-004-139B-1

  Query Match             1.6%;  Score 34;  DB 1;  Length 5837;
  Best Local Similarity   57.5%;  Pred. No. 4.1;
  Matches   61;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy   1405 GTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAG 1464
            || |||| | |||  |         | ||||     || ||||||  ||| | |||| |||
Db   3686 GTCATCGTCCAGAGTTAGTGCTTCAACACCTTCGAGAATGCAGTATTCTTTTTCGCCAAT 3627

Qy   1465 AGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCA 1510
             ||| || |     ||| || ||| ||| |     | | ||| || |
Db   3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTCACAAATCTA 3581


RESULT 13
US-08-117-491-1/c
; Sequence 1, Application US/08117491
; Patent No. 5500363
;  GENERAL INFORMATION:
;     APPLICANT:  Comb, Donald G.
;     APPLICANT:  Perler, Francine
;     APPLICANT:  Kucera, Rebecca
;     APPLICANT:  Jack, William E.
;     TITLE OF INVENTION:  RECOMBINANT THERMOSTABLE DNA
;     TITLE OF INVENTION:  POLYMERS FROM ARCHAEBACTERIA
;     NUMBER OF SEQUENCES:  25
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS
;       ADDRESSEE:  & CUSHMAN
;       STREET:  130 WATER STREET
;       CITY:  BOSTON
;       STATE:  MASSACHUSETTS
;       COUNTRY:  US
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/117,491
;       FILING DATE:  07-SEP-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/811,421
;       FILING DATE:  12-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/686,340
;       FILING DATE:  17-APR-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/626,057
;       FILING DATE:  11-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/513,994
;       FILING DATE:  26-APR-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Resnick, David S.
;       REGISTRATION NUMBER:  34,235
;       REFERENCE/DOCKET NUMBER:  39296C3FWC2
;     TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  (617) 523-3400
;       TELEFAX:  (617) 523-6440
;       TELEX:  200291 STRE UR
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5837 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-117-491-1

  Query Match             1.6%;  Score 34;  DB 1;  Length 5837;
  Best Local Similarity   57.5%;  Pred. No. 4.1;
  Matches   61;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy    1405 GTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAG 1464
           || |||| | |||  |         | ||||      || ||||||  ||| | |||| |||
Db    3686 GTCATCGTCCAGAGTTAGTGCTTCAACACCTTCGAGAATGCAGTATTCTTTTTCGCCAAT 3627

Qy    1465 AGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCA 1510
             ||| || |     ||| || ||| ||| |     | | ||| || |
Db    3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTTCACAAATCTA 3581


RESULT 14
US-08-271-364A-1/c
; Sequence 1, Application US/08271364A
; Patent No. 5756334
;  GENERAL INFORMATION:
;     APPLICANT:  PERLER, FRANCINE B.
;     APPLICANT:  SOUTHWORTH, MAURICE W.
;     TITLE OF INVENTION:  RECOMBINANT THEROMSTABLE DNA POLYMERASE
;     TITLE OF INVENTION:  FROM ARCHAEBACTERIA
;     NUMBER OF SEQUENCES:  42
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  NEW ENGLAND BIOLABS, INC.
;       STREET:  32 TOZER ROAD
;       CITY:  BEVERLY
;       STATE:  MASSACHUSETTS
;       COUNTRY:  US
;       ZIP:  01915
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/271,364A
;       FILING DATE:  06-JUL-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/811,421
;       FILING DATE:  18-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/686,340
;       FILING DATE:  17-APR-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/626,057
;       FILING DATE:  11-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/513,994
;       FILING DATE:  26-APR-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  WILLIAMS, GREGORY D.
;       REGISTRATION NUMBER:  30901
;       REFERENCE/DOCKET NUMBER:  NEB-101
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (508) 927-5054
;       TELEFAX:  (508) 927-1705
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5837 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  not relevant
US-08-271-364A-1

  Query Match             1.6%;  Score 34;  DB 1;  Length 5837;
  Best Local Similarity   57.5%;  Pred. No. 4.1;
  Matches   61;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy    1405 GTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAG 1464
           || |||| | |||  |         | ||||      || ||||||  ||| | |||| |||
Db    3686 GTCATCGTCCAGAGTTAGTGCTTCAACACCTTCGAGAATGCAGTATTCTTTTTCGCCAAT 3627

Qy    1465 AGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCA 1510
             ||| || |     ||| || ||| ||| |     | | ||| || |
Db    3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTTCACAAATCTA 3581


RESULT 15
US-08-811-492-1/c
; Sequence 1, Application US/08811492
; Patent No. 5834247
;  GENERAL INFORMATION:
;     APPLICANT:  COMB, DONALD G.
;     APPLICANT:  PERLER, FRANCINE B.
;     APPLICANT:  JACK, WILLIAM E.
;     APPLICANT:  XU, MING-QUN
;     APPLICANT:  HODGES, ROBERT A.
;     APPLICANT:  NOREN, CHRISTOPHER J.
;     APPLICANT:  CHONG, SHAORONG S.C.
;     APPLICANT:  ADAM, ERIC
;     APPLICANT:  SOUTHWORTH, MAURICE
;     TITLE OF INVENTION:  MODIFIED PROTEINS, METHODS OF THEIR
;     TITLE OF INVENTION:  PRODUCTION AND METHODS FOR PURIFICATION OF TARGET
;     TITLE OF INVENTION:  PROTEINS
;     NUMBER OF SEQUENCES:  155
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  GREGORY D. WILLIAMS; NEW ENGLAND BIOLABS, INC.
;       STREET:  32 TOZER RAOD
;       CITY:  BEVERLY
;       STATE:  MASSACHUSETTS
;       COUNTRY:  USA
;       ZIP:  01915
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC\_DOS/MS\_DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/811,492
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/580,555
;       FILING DATE:  29-DEC-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/496,247
;       FILING DATE:  28-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/146,885
;       FILING DATE:  03-NOV-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/004,139
;       FILING DATE:  09-DEC-1992
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Williams, Gregory D
;       REGISTRATION NUMBER:  30901
;       REFERENCE/DOCKET NUMBER:  NEB-036C4
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  508-927-5054
```

```
;      TELEFAX:  509-927-1705
;      TELEX:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;      LENGTH:  5837 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
US-08-811-492-1

  Query Match             1.6%;  Score 34;  DB 2;  Length 5837;
  Best Local Similarity   57.5%;  Pred. No. 4.1;
  Matches   61;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy   1405 GTAATCGACAAGAAATTCAAAGGCGACACGAAGAAGTATGCAGACTTCGTATTCGACAAG 1464
          || |||| | |||  |       | ||||     || ||||||  |||   | |||| |||
Db   3686 GTCATCGTCCAGAGTTAGTGCTTCAACACCTTCGAGAATGCAGTATTCTTTTTCGCCAAT 3627

Qy   1465 AGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCA 1510
            ||| || |    ||| || ||| ||| |    | | || | || |
Db   3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTTCACAAATCTA 3581

Search completed: May 23, 2003, 04:44:48
Job time : 122 secs
```

D
D

```
                         GenCore version 5.1.4_p5_4578
                Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         May 23, 2003, 03:33:01 ; Search time 529 Seconds
                                          (without alignments)
                                          5339.261 Million cell updates/sec

Title:          US-10-008-355-1
Perfect score:  2139
Sequence:       1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters:      1657494

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                1:  /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                10:  /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                12:  /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
                13:  /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                14:  /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2139	100.0	2139	9	US-10-008-355-1	Sequence 1, Appli
	2	40.8	1.9	867	9	US-10-123-155-20	Sequence 20, Appl
	3	36.2	1.7	345	9	US-09-738-626-2081	Sequence 2081, Ap
	4	36.2	1.7	2281	10	US-09-900-497-1	Sequence 1, Appli
	5	36	1.7	292	10	US-09-294-093B-856	Sequence 856, App
	6	36	1.7	369	10	US-09-878-574-5240	Sequence 5240, Ap
c	7	35.2	1.6	671	9	US-10-184-644-346	Sequence 346, App
c	8	35.2	1.6	671	9	US-10-184-634-346	Sequence 346, App
	9	35	1.6	1896	9	US-10-213-990-53	Sequence 53, Appl
	10	35	1.6	2060	9	US-10-213-990-52	Sequence 52, Appl
	11	34.4	1.6	489	9	US-10-184-644-116	Sequence 116, App
	12	34.4	1.6	489	9	US-10-184-634-116	Sequence 116, App
	13	34	1.6	268	10	US-09-923-876-5908	Sequence 5908, Ap
c	14	34	1.6	50000	9	US-10-152-724A-22	Sequence 22, Appl
	15	33.6	1.6	50937	9	US-09-808-880-1	Sequence 1, Appli
c	16	33.4	1.6	802	9	US-10-184-644-312	Sequence 312, App
c	17	33.4	1.6	802	9	US-10-184-634-312	Sequence 312, App
	18	33.2	1.6	256	10	US-09-878-574-9837	Sequence 9837, Ap
	19	33.2	1.6	607	9	US-10-123-155-344	Sequence 344, App
	20	33	1.5	999	9	US-10-184-644-434	Sequence 434, App
	21	33	1.5	999	9	US-10-184-634-434	Sequence 434, App
c	22	32.4	1.5	171	9	US-09-754-853A-273	Sequence 273, App
c	23	32.4	1.5	520	9	US-10-184-644-332	Sequence 332, App
c	24	32.4	1.5	520	9	US-10-184-634-332	Sequence 332, App
c	25	32.4	1.5	777	10	US-09-910-943-568	Sequence 568, App
	26	32.4	1.5	846	9	US-09-738-626-1515	Sequence 1515, Ap
c	27	32.4	1.5	987	10	US-09-974-300-1183	Sequence 1183, Ap
c	28	32.4	1.5	1011	10	US-09-974-300-2041	Sequence 2041, Ap
c	29	32.4	1.5	513509	9	US-09-754-853A-4	Sequence 4, Appli
	30	32.2	1.5	671	9	US-10-184-644-346	Sequence 346, App
	31	32.2	1.5	671	9	US-10-184-634-346	Sequence 346, App
	32	32.2	1.5	860	10	US-09-770-445-596	Sequence 596, App
	33	32.2	1.5	2010	12	US-10-032-717-9	Sequence 9, Appli
	34	32.2	1.5	2169	9	US-10-108-580-1	Sequence 1, Appli
	35	32.2	1.5	3468	9	US-09-988-462-2	Sequence 2, Appli
c	36	32	1.5	418	9	US-09-918-995-6839	Sequence 6839, Ap
	37	32	1.5	491	9	US-10-123-155-278	Sequence 278, App
	38	32	1.5	1941	9	US-10-090-455-3	Sequence 3, Appli
	39	32	1.5	1947	9	US-09-988-462-3	Sequence 3, Appli
	40	32	1.5	3455	9	US-10-072-621-4	Sequence 4, Appli
	41	32	1.5	3455	9	US-10-090-455-1	Sequence 1, Appli
	42	32	1.5	3468	9	US-09-988-462-4	Sequence 4, Appli
	43	32	1.5	3468	9	US-09-988-462-8	Sequence 8, Appli
	44	32	1.5	3546	9	US-09-988-462-10	Sequence 10, Appl
	45	32	1.5	3546	9	US-09-988-462-12	Sequence 12, Appl

ALIGNMENTS

```
RESULT 1
US-10-008-355-1
; Sequence 1, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 2139
;   TYPE: DNA
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-1

  Query Match             100.0%;  Score 2139;  DB 9;  Length 2139;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2139;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60

Qy         61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120

Qy        121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180

Qy        181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240
```

```
Qy    241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300

Qy    301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360

Qy    361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420

Qy    421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480

Qy    481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540

Qy    541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600

Qy    601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660

Qy    661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720

Qy    721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780

Qy    781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840

Qy    841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900

Qy    901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960

Qy    961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020

Qy   1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080

Qy   1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140

Qy   1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200

Qy   1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260

Qy   1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320

Qy   1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380

Qy   1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440

Qy   1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500

Qy   1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560

Qy   1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620

Qy   1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680

Qy   1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740

Qy   1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800

Qy   1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860

Qy   1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920

Qy   1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980

Qy   1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040

Qy   2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100

Qy   2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 2139
          |||||||||||||||||||||||||||||||||||||||
Db   2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATCTAA 2139
```

```
RESULT 2
US-10-123-155-20
; Sequence 20, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
```

```
; APPLICANT:  Stewart,Timothy A.
; APPLICANT:  Tumas,Daniel
; APPLICANT:  Watanabe,Colin K
; APPLICANT:  Wood,William
; APPLICANT:  Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION:  ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE:  2002-04-15
; Prior Application removed -  See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 20
;  LENGTH: 867
;  TYPE: PRT
;  ORGANISM: Homo Sapien
US-10-123-155-20

  Query Match             1.9%;  Score 40.8;  DB 9;  Length 867;
  Best Local Similarity   7.6%;  Pred. No. 0.01;
  Matches   55;  Conservative  237;  Mismatches  427;  Indels    1;  Gaps    1;

Qy    691 GTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAG 750
Db     33 GRFQRDRRNIRPNIILVLTDDQDVELGSMQVMNKTRRIMEQGGAHFINAFVTTPMCCPSR 92

Qy    751 CCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACC 810
Db     93 SSILTGKYVHNHNTYTNNENCSSPSWQAQHESRTFAVYLNSTGYRTAFFGKYLNEYNGSY 152

Qy    811 ATCGGTTTCCCGGGCAGTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATC 870
Db    153 VPPGWKEWVGLLKNSRFYNYTLCRNGVKEKHGSDYSKDYLTDLITNDSVSFFRTSKKMYP 212

Qy    871 GAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAA 930
Db    213 HRPVLMVISHAAPHGPEDSAPQYSRLFPNASQHITPSYNYAPNPDKHWIMRYTGPMKPIH 272

Qy    931 GCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCT 990
Db    273 MEFTNMLQRKRLQTLMSVDDSMETIYNMLVETGELDNTYIVYTADHGYHIGQFGLVKGKS 332

Qy    991 AACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGT 1050
Db    333 MPYEFDIRVPFYVRGPNVEAGCLNPHIVLNIDLAPTILDIAGLDIPADMDGKSILKLLDT 392

Qy   1051 CGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCT 1110
Db    393 ERPVNRFHLKKKMRVWRDSFLVERGKLLHKRDNDKVDAQEENFLPKYQRVKDLCQRAEYQ 452

Qy   1111 GTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAAC 1170
Db    453 TACEQLGQKWQCVEDATGKLKLHKCK-GPMRLGGSRALSNLVPKYYGQGSEACTCDSGDY 511

Qy   1171 CGTGAGATGACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCA 1230
Db    512 KLSLAGRRKKLFKKKYKASYVRSRSIRSVAIEVDGRVYHVGLGDAAQPRNLTKRHWPGAP 571

Qy   1231 CAGTTTGCCAACGCATTGGCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTT 1290
Db    572 EDQDDKDGGDFSGTGGLPDYSAANPIKVTHRCYILENDTVQCDLDLYKSLQAWKDHKLHI 631

Qy   1291 GACGACAAGTACAAAGACTACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTC 1350
Db    632 DHEIETLQNKIKNLREVRGHLKKKRPEECDCHKISYHTQHKGRLKHRGSSLHPFRKGLQE 691

Qy   1351 GATATTGTACGCCGGCGTATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATC 1410
Db    692 KDKVWLLREQKRKKKLRKLLKRLQNNDTCSMPGLTCFTHDNQHWQTAPFWTLGPFCACTS 751

RESULT 3
US-09-738-626-2081
; Sequence 2081, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2081
;   LENGTH: 345
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-2081

  Query Match             1.7%;  Score 36.2;  DB 9;  Length 345;
  Best Local Similarity   57.5%;  Pred. No. 0.21;
  Matches   65;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy    521 TCTGCATCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTAT 580
Db    233 TCGGCATCTTCAATTTCAAATATGCCAACAACAAATACTTCGTCGTAGGAATCAATGACA 292

Qy    581 TCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGA 633
Db    293 TCACTGGTCTTGGTCTTGGATCCTCTGATCAGAATCCCCTAGGAACCTTCGGA 345

RESULT 4
US-09-900-497-1
; Sequence 1, Application US/09900497
; Patent No. US20020082232A1
; GENERAL INFORMATION:
;  APPLICANT: Allen, Keith D.
;  TITLE OF INVENTION: TRANSGENIC MICE CONTAINING NPY6-R
;  TITLE OF INVENTION:  NEUROPEPTIDE RECEPTOR GENE DISRUPTIONS
;  FILE REFERENCE: R-639
;  CURRENT APPLICATION NUMBER: US/09/900,497
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US 60/216,260
;  PRIOR FILING DATE: 2000-07-06
;  PRIOR APPLICATION NUMBER: US 60/221,474
;  PRIOR FILING DATE: 2000-07-27
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2281
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-900-497-1

  Query Match             1.7%;  Score 36.2;  DB 10;  Length 2281;
  Best Local Similarity   55.0%;  Pred. No. 0.65;
  Matches   71;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy    409 GACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCT 468
```

```
Db     502 GACAAGGAGGATCAGAAGCTTAAGGACATCATTTTGTACATAGTGAGTTTGAGGAAAGCT 561

Qy     469 CAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATC 528
           |||    |||  |     ||  |    ||| | ||  || |  | ||| ||  || |  |
Db     562 GAGGTTACATGGAACTCTCTCTCTCTCAAAAACAAAACAAAACAAAACAAAACCTTCTAC 621

Qy     529 GTAGAGCCT 537
            .  | |  ||
Db     622 TAATATTCT 630

RESULT 5
US-09-294-093B-856
; Sequence 856, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
;  APPLICANT:      Lalgudi, Raghunath, V.
;  APPLICANT:  Ito, Laura, Y.
;  APPLICANT:  Sherman, Bradley, K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
;  FILE REFERENCE:  PL-0009 US
;  CURRENT APPLICATION NUMBER: US/09/294,093B
;  CURRENT FILING DATE:  1999-04-16
;  PRIOR APPLICATION NUMBER:  60/082,567
;  PRIOR FILING DATE:  April 21, 1998
;  NUMBER OF SEQ ID NOS:      6207
;  SOFTWARE:       PERL Program
; SEQ ID NO 856
;   LENGTH: 292
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20010051335A1 700343414H1
;   NAME/KEY: unsure
;   LOCATION: 111
;   OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-856

  Query Match             1.7%;  Score 36;  DB 10;  Length 292;
  Best Local Similarity   51.9%;  Pred. No. 0.22;
  Matches   81;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy    1873 GGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACG 1932
           || | |||  |||||| |||  ||        |  | | ||||||  ||  || ||  ||||
Db     134 GGGCTCTACTCCGAGATCGGGAAGAAGGCCAGAGATCTCCTATATAAGGACTACCACACG 193

Qy    1933 GGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTC 1992
           | |   || | |     | ||   | ||    |||||   |    ||    |  |||    |
Db     194 GACCAGAAGTTCACGCTCACCACCTACGCCGCCAACGGAGCTGCCATTACTGCTGCAAGC 253

Qy    1993 GATGGCAACTGGGAAGCTATGAGTGGTGACATCGAG 2028
               | ||    ||||||||    |  ||| ||| ||
Db     254 ACGAGGAAAGATGAAGCTATCTTTAATGAAATCCAG 289

RESULT 6
US-09-878-574-5240
; Sequence 5240, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 5240
;   LENGTH: 369
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: LIB3028-035-Q1-B1-H8
US-09-878-574-5240

  Query Match             1.7%;  Score 36;  DB 10;  Length 369;
  Best Local Similarity   55.6%;  Pred. No. 0.25;
  Matches   69;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy    1513 GACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGC 1572
           ||  | ||| |  |||         |   |  ||||| || |||| |||    ||||||||
Db     142 GAGGAAGAAGAAATTGAGGTTCCAGTTTACAAAGAACCTGCAGGAGAAAATTCCAAGATG 201

Qy    1573 GTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAG 1632
           | ||  | ||  ||||   |||   |    || | |||     || | | |       ||| |||
Db     202 GAAACTGATGAAGCTCCTGCTGATGCTGCTGCTGCTGCAACTCCTAGCACCAATGAC 261

Qy    1633 AAGG 1636
            . ||  |
Db     262 AATG 265

RESULT 7
US-10-184-644-346/c
; Sequence 346, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C227
;  CURRENT APPLICATION NUMBER: US/10/184,644
;  CURRENT FILING DATE:  2002-06-28
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 346
;   LENGTH: 671
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-184-644-346

  Query Match             1.6%;  Score 35.2;  DB 9;  Length 671;
  Best Local Similarity   7.1%;  Pred. No. 0.67;
  Matches   29;  Conservative  138;  Mismatches  244;  Indels    0;  Gaps    0;

Qy    1577 TAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGG 1636
           |  :  :::| :      ::::: : :  |:    :::   : :: |: |:         :
Db     660 TY.Y..YHSC.YH.H.C.YBSHD.H.SSCGY.C..SHYC..D.MH.TS.CM.SAT.CNB. 601

Qy    1637 GCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCG 1696
            : : :  : |    |  |      | :    : : :  : : :  :      | ::::  |::
Db     600 TS.HAS.MH.TM...T..CT...CT.M..TM.M.KCTMNBM.BT..M.C.BYSM.YARB. 541

Qy    1697 ATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACG 1756
            |     :: |::::      | : | :    : :  |           :: ::::   :
Db     540 TTM.KBHMNTMRYY.MM.ATBBHANM....SAB..AC....TN.GYMBH.SHBM.T.M.. 481

Qy    1757 GTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGA 1816
           :: :       :   : : : :::: : | : : |   ::: :|        :::      ::   |
Db     480 SD.M..M..BS..MMM.BMYHKS.C.SYMYAY.BMB.MT..M...YBYM.M.C.MYT..A 421

Qy    1817 GCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTC 1876
```

```
               :::    :         :               |: : ::|:: :    : :   :    |     |
Db     420 SSS..MMTSMMT.Y...T...SHSSS..CYCYCMYCYMMB..HH.Y..HCSSCCSCY.T. 361

Qy    1877 GCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCG 1936
            :: : ::      :    :    :      :: :|    :: |     :: :   :  : :: :::
Db     360 CYHTY.YYY.M..M...Y.YY......YYYYT...YYYT.HYYMW.S.YHB.HSHSHSSS 301

Qy    1937 GTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTG 1987
              :   :   :   :    :      |  :   :   : :: ::: : :   :::|::|:
Db     300 SS..Y..M.MYCY.M...M.T.MYCY..MMBSBHSHSSSSSSSSSSTSYTK 250


RESULT 8
US-10-184-634-346/c
; Sequence 346, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C217
;  CURRENT APPLICATION NUMBER: US/10/184,634
;  CURRENT FILING DATE:  2002-06-28
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 346
;   LENGTH: 671
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-184-634-346

  Query Match             1.6%;  Score 35.2;  DB 9;  Length 671;
  Best Local Similarity   7.1%;  Pred. No. 0.67;
  Matches   29;  Conservative  138;  Mismatches  244;  Indels    0;  Gaps    0;

Qy    1577 TAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGG 1636
            |  :  :::| :       ::::: : :  |:     :::    : ::  |: |:         :
Db     660 TY.Y..YHSC.YH.H.C.YBSHD.H.SSCGY.C..SHYC..D.MH.TS.CM.SAT.CNB. 601

Qy    1637 GCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCGAGCG 1696
             : : :   : |      |   |     | :    :  :  :  : : :     | ::::  ]::
Db     600 TS.HAS.MH.TM...T..CT...CT.M..TM.M.KCTMNBM.BT..M.C.BYSM.YARB. 541

Qy    1697 ATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACG 1756
             |     :: |::::     | : |  :     :  :   |             :: ::::    :
Db     540 TTM.KBHMNTMRYY.MM.ATBBHANM....SAB..AC....TN.GYMBH.SHBM.T.M.. 481

Qy    1757 GTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAGA 1816
            :: :       :  : : :::: :  | : : |   ::: :|          :::       ::    |
Db     480 SD.M..M..BS..MMM.BMYHKS.C.SYMYAY.BMB.MT..M...YBYM.M.C.MYT..A 421

Qy    1817 GCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGTC 1876
            :::    :         :               |: : ::|:: :    : :   :    |     |
Db     420 SSS..MMTSMMT.Y...T...SHSSS..CYCYCMYCYMMB..HH.Y..HCSSCCSCY.T. 361

Qy    1877 GCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACGGGCG 1936
            :: : ::      :    :    :      :: :|    :: |     :: :   :  : :: :::
Db     360 CYHTY.YYY.M..M...Y.YY......YYYYT...YYYT.HYYMW.S.YHB.HSHSHSSS 301

Qy    1937 GTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTG 1987
              :   :   :   :    :      |  :   :   : :: ::: : :   :::|::|:
Db     300 SS..Y..M.MYCY.M...M.T.MYCY..MMBSBHSHSSSSSSSSSSTSYTK 250
```

```
RESULT 9
US-10-213-990-53
; Sequence 53, Application US/10213990
; Publication No. US20030082595A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Bussey, Howard
;  APPLICANT:  Storms, Reg
;  APPLICANT:  Roemer, Terry
;  TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUMIGATUS ENCODING INDUSTRIAL
;  TITLE OF INVENTION:  ENZYMES AND METHODS OF USE
;  FILE REFERENCE: 10182-019-999
;  CURRENT APPLICATION NUMBER: US/10/213,990
;  CURRENT FILING DATE:  2002-08-05
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 53
;   LENGTH: 1896
;   TYPE: DNA
;   ORGANISM: Aspergillus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1896)
US-10-213-990-53

  Query Match             1.6%;  Score 35;  DB 9;  Length 1896;
  Best Local Similarity   48.3%;  Pred. No. 1.5;
  Matches   98;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy     645 CAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGC 704
            | | | | |   |  ||     || ||  | ||| | || |      |   |     ||||| | |
Db     927 CGATTCGTTCCGCTCGGTCTACGCGATCAACTCCGGCATCCCACAGGGCGCTGCCGTTTC 986

Qy     705 CGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGC 764
            ||    |||      ||||  ||   |  |  ||||        |   |    |   |  |     |  |
Db     987 CGCTGGCCGCTACCCCGAGGACGTCTACTACAACGGCAACCCTTGGTTCCTCACCACCCT 1046

Qy     765 TGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGG 824
             ||||    ||   |||    |||| | |||   | || |      | ||  ||||||| |      |
Db    1047 CGCCGCTGCCGAGCAGCTCTACGACGCTATCTACCAGTGGAAGAAGATCGGTTCCATCAG 1106

Qy     825 CAGTACGGATCGCTACCTCACTT 847
            ||  ||     .   || |||| | |
Db    1107 CATCACCAGCACCTCCCTCGCCT 1129


RESULT 10
US-10-213-990-52
; Sequence 52, Application US/10213990
; Publication No. US20030082595A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Bo
;  APPLICANT:  Bussey, Howard
;  APPLICANT:  Storms, Reg
;  APPLICANT:  Roemer, Terry
;  TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUMIGATUS ENCODING INDUSTRIAL
;  TITLE OF INVENTION:  ENZYMES AND METHODS OF USE
;  FILE REFERENCE: 10182-019-999
;  CURRENT APPLICATION NUMBER: US/10/213,990
;  CURRENT FILING DATE:  2002-08-05
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 52
;   LENGTH: 2060
;   TYPE: DNA
;   ORGANISM: Aspergillus
US-10-213-990-52

  Query Match             1.6%;  Score 35;  DB 9;  Length 2060;
  Best Local Similarity   48.3%;  Pred. No. 1.5;
  Matches   98;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;
```

```
Qy      645 CAACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGC 704
            | | | | |  |  ||     || ||  | ||| | || |    |  |    ||||| | |
Db     1091 CGATTCGTTCCGCTCGGTCTACGCGATCAACTCCGGCATCCCACAGGGCGCTGCCGTTTC 1150

Qy      705 CGACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGC 764
            || .  |||       ||||  ||  |  |  ||||       |   |   |  |    |   |
Db     1151 CGCTGGCCGCTACCCCGAGGACGTCTACTACAACGGCAACCCTTGGTTCCTCACCACCCT 1210

Qy      765 TGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGG 824
             ||||  ||  |||   |||| | |||   | || |     | ||  ||||||| |     |
Db     1211 CGCCGCTGCCGAGCAGCTCTACGACGCTATCTACCAGTGGAAGAAGATCGGTTCCATCAG 1270

Qy      825 CAGTACGGATCGCTACCTCACTT 847
            || ||      || |||| | |
Db     1271 CATCACCAGCACCTCCCTCGCCT 1293
```

RESULT 11
US-10-184-644-116
; Sequence 116, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 116
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-116

```
  Query Match             1.6%;  Score 34.4;  DB 9;  Length 489;
  Best Local Similarity   8.3%;  Pred. No. 1;
  Matches   37;  Conservative  116;  Mismatches  291;  Indels    0;  Gaps    0;

Qy       66 CAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAAT 125
             : | :: : :::     : : : |  :|         |:      :   : |   | ::: :|
Db        1 MEAPDYEVLSVREQLFHERIRECIISTLLFATLYILCHIFLTRFKKPAEFTTVDDEDATV 60

Qy      126 GCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCAT 185
                     ||          :      :   :       :     :   :   :   :
Db       61 NKIALELCTFTLAIALGAVLLLPFSIISNEVLLSLPRNYYIQWLNGSLIHGLWNLVFLFP 120

Qy      186 TGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGG 245
                :          :   |          :::|: |::: |:::      |   :  |  ::: :|  ::: :
Db      121 NLSLIFLMPFAYFFTESEGFAGSRKGVLGRVYETVVMLMLLTLLVLGMVWVASAIVDKNK 180

Qy      246 CCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGA 305
                :: : ::  :   :| :    :             |  :  |     :          :
Db      181 ANRESLYDFWEYYLPYLYSCISFLGVLLLLVCTPLGLARMFSVTGKLLVKPRLLEDLEEQ 240

Qy      306 TCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCC 365
             : :       |  |:  |  |:  :    :     :     :    | ::     ::::        :    :
Db      241 LYCSAFEEAALTRRICNPTSCWLPLDMELLHRQVLALQTQRVLLEKRRKASAWQRNLGYP 300

Qy      366 GGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACA 425
                     |      | ::    | :          : | :   |: |         :: :   |:    :   ::
Db      301 LAMLCLLVLTGLSVLIVAIHILELLIDEAAMPRGMQGTSLGQVSFSKLGSFGAVIQVVLI 360

Qy      426 GCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGA 485
              : :: :::   ::       :: : :: :   :          |    :       :  :|    :  :      |   :  :
Db      361 FYLMVSSVVGFYSSPLFRSLRPRWHDTAMTQIIGNCVCLLVLSSALPVFSRTLGLTRFDL 420

Qy      486 ACTGGCCAAAAAAGAAAATGCAGA 509
                  : |     :          :   :   | |
Db      421 LGDFGRFNWLGNFYIVFLYNAAFA 444
```

RESULT 12
US-10-184-634-116
; Sequence 116, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 116
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-116

```
  Query Match             1.6%;  Score 34.4;  DB 9;  Length 489;
  Best Local Similarity   8.3%;  Pred. No. 1;
  Matches   37;  Conservative  116;  Mismatches  291;  Indels    0;  Gaps    0;

Qy       66 CAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAAT 125
Db        1 MEAPDYEVLSVREQLFHERIRECIISTLLFATLYILCHIFLTRFKKPAEFTTVDDEDATV 60

Qy      126 GCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCAT 185
Db       61 NKIALELCTFTLAIALGAVLLLPFSIISNEVLLSLPRNYYIQWLNGSLIHGLWNLVFLFP 120

Qy      186 TGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGG 245
Db      121 NLSLIFLMPFAYFFTESEGFAGSRKGVLGRVYETVVMLMLLTLLVLGMVWVASAIVDKNK 180

Qy      246 CCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGA 305
Db      181 ANRESLYDFWEYYLPYLYSCISFLGVLLLLVCTPLGLARMFSVTGKLLVKPRLLEDLEEQ 240

Qy      306 TCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCC 365
Db      241 LYCSAFEEAALTRRICNPTSCWLPLDMELLHRQVLALQTQRVLLEKRRKASAWQRNLGYP 300

Qy      366 GGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAGGACA 425
Db      301 LAMLCLLVLTGLSVLIVAIHILELLIDEAAMPRGMQGTSLGQVSFSKLGSFGAVIQVVLI 360

Qy      426 GCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGA 485
Db      361 FYLMVSSVVGFYSSPLFRSLRPRWHDTAMTQIIGNCVCLLVLSSALPVFSRTLGLTRFDL 420

Qy      486 ACTGGCCAAAAAAGAAAATGCAGA 509
```

```
Db    421 LGDFGRFNWLGNFYIVFLYNAAFA 444

RESULT 13
US-09-923-876-5908
; Sequence 5908, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/923,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 5908
;   LENGTH: 268
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700457955H1
;   NAME/KEY: unsure
;   LOCATION: 13, 50, 69, 179
;   OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-5908

  Query Match             1.6%;  Score 34;  DB 10;  Length 268;
  Best Local Similarity   48.4%;  Pred. No. 0.98;
  Matches   91;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy    436 ATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAA 495

Db     63 ATCACTNCGCAACCATGGCGGCGGCAGCAGGCGGAGGCGCGGTGCGTGCTACTGTCCCAG 122

Qy    496 AAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCAACAACGAA 555

Db    123 AAGGAGCAGGACATACAGATGATGCTCGCCGCCGACGTCCACCTTGGCACCAAGAANTGC 182

Qy    556 TACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTATTTGCTCCTCCCAGC 615

Db    183 GATTTCCAGACGGAGCGCTATGTCTTTAAGCGCGTTCCGACGGCATCTACATCATCAATC 242

Qy    616 TCTGTAGG 623

Db    243 TTGGCAAG 250

RESULT 14
US-10-152-724A-22/c
; Sequence 22, Application US/10152724A
; Publication No. US20030082714A1
; GENERAL INFORMATION:
;  APPLICANT: LITTLE, Melissa
;  APPLICANT:  HOLMES, Gregory
;  APPLICANT:  KOLLE, Gabriel
;  APPLICANT:  YAMADA, Toshiya
;  APPLICANT:  GEORGAS, Kylie
;  APPLICANT:  WILKINSON, Lorine
;  TITLE OF INVENTION: No. US20030082714A1el Nucleic Acid and Polypeptide
;  FILE REFERENCE: P22378
;  CURRENT APPLICATION NUMBER: US/10/152,724A
;  CURRENT FILING DATE:  2002-05-23
;  PRIOR APPLICATION NUMBER: Australian App No. US20030082714A1PQ 4348
;  PRIOR FILING DATE: 1999-11-26
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-152-724A-22

  Query Match             1.6%;  Score 34;  DB 9;  Length 50000;
  Best Local Similarity   56.1%;  Pred. No. 23;
  Matches   64;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    362 TTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAAG 421

Db  18679 TTCTGAAACGTTTGATGAGAGATCAGAGCCAGAGCATGAAAGAGACAGAGAGGAAAAAAG 18620

Qy    422 GACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGG 475

Db  18619 GCCAGCAGAACTTTACCGCTGAAATGACAAGGATGCTGAGCTCTGTTGAGGTGG 18566

RESULT 15
US-09-808-880-1
; Sequence 1, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 50937
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-808-880-1

  Query Match             1.6%;  Score 33.6;  DB 9;  Length 50937;
  Best Local Similarity   47.2%;  Pred. No. 31;
  Matches  102;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy   1377 CGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAA 1436

Db  44477 CGCCGAGATCTACGACGCCGTGTACCGGGGCCGGGGCAAGGACTACGCCGGCGAGGCGAA 44536

Qy   1437 GAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGC 1496

Db  44537 GGACGTGGCGGACCTCGTGCGCGACCGGGTGCCGGACGCGTCCTCCCTCCTGGACGTGGC 44596

Qy   1497 CATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGT 1556

Db  44597 CTGCGGCACGGGCGCGCACCTGCGGCACTTCGCCACGCTCTTCGACGACGCCCGCGGTCT 44656

Qy   1557 AGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGC 1592

Db  44657 CGAACTGTCCGCGAGCATGCTGGACATCGCCCGCTC 44692

Search completed: May 23, 2003, 07:01:24
Job time : 741 secs,
```

D D


```
OM nucleic - nucleic search, using sw model

Run on:          May 23, 2003, 02:45:01 ; Search time 2048 Seconds
                                          (without alignments)
                                          16915.110 Million cell updates/sec

Title:           US-10-008-355-1
Perfect score:   2139
Sequence:        1 atgcaaatgaaattaaaaag..........aagagctgaagttgatctaa 2139

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters:      32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                 1:  em_estba:*
                 2:  em_esthum:*
                 3:  em_estin:*
                 4:  em_estmu:*
                 5:  em_estov:*
                 6:  em_estpl:*
                 7:  em_estro:*
                 8:  em_htc:*
                 9:  gb_est1:*
                 10:  gb_est2:*
                 11:  gb_htc:*
                 12:  gb_est3:*
                 13:  gb_est4:*
                 14:  gb_est5:*
                 15:  em_estfun:*
                 16:  em_estom:*
                 17:  gb_gss:*
                 18:  em_gss_hum:*
                 19:  em_gss_inv:*
                 20:  em_gss_pln:*
                 21:  em_gss_vrt:*
                 22:  em_gss_fun:*
                 23:  em_gss_mam:*
                 24:  em_gss_mus:*
                 25:  em_gss_other:*
                 26:  em_gss_pro:*
                 27:  em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	110	5.1	537	17	BH388186	BH388186 AG-ND-124
c	2	110	5.1	591	17	BH392840	BH392840 AG-ND-162
c	3	108.2	5.1	500	17	BH386991	BH386991 AG-ND-148
	4	103.8	4.9	418	17	BH375201	BH375201 AG-ND-173
c	5	72.4	3.4	426	17	BH400391	BH400391 AG-ND-147
c	6	66.4	3.1	697	17	BH375688	BH375688 AG-ND-120
	7	55	2.6	398	10	AW774807	AW774807 EST333958
c	8	49.2	2.3	1005	17	BH164253	BH164253 ENTTG63TF
c	9	42.8	2.0	555	17	BH400866	BH400866 AG-ND-125
	10	42.2	2.0	692	9	AI405287	AI405287 GH25242.5
	11	41.2	1.9	721	17	BH391233	BH391233 AG-ND-141
	12	39.4	1.8	822	17	BH371846	BH371846 AG-ND-162
	13	39	1.8	606	9	AL728310	AL728310 AL728310
c	14	38.6	1.8	687	17	BH448937	BH448937 BOGZL67TR
	15	38.4	1.8	184	10	AW375157	AW375157 QV2-CT012
	16	38.4	1.8	466	12	BF426788	BF426788 df72a12.y
	17	38.4	1.8	491	10	AW148212	AW148212 da13h05.x
	18	38.4	1.8	505	14	BQ163407	BQ163407 952078A08
	19	38.4	1.8	540	14	BQ035218	BQ035218 952078A08
	20	38.4	1.8	600	14	BQ163454	BQ163454 952078F07
c	21	38.4	1.8	895	12	BF784434	BF784434 602110820
	22	38	1.8	465	10	AV628406	AV628406 AV628406
	23	38	1.8	497	10	AV621707	AV621707 AV621707
	24	38	1.8	587	13	BI247683	BI247683 602959191
c	25	38	1.8	1091	10	BE282103	BE282103 601100927
	26	37.8	1.8	422	10	BB851067	BB851067 BB851067
	27	37.8	1.8	455	14	BQ241220	BQ241220 TaE05007G
c	28	37.8	1.8	461	9	AL372451	AL372451 MtBA51A03
c	29	37.8	1.8	680	13	BI309279	BI309279 EST530689
c	30	37.8	1.8	773	12	BG644863	BG644863 EST506482
	31	37.6	1.8	422	10	AV629946	AV629946 AV629946
	32	37.6	1.8	927	12	BG437679	BG437679 602489932
	33	37.4	1.7	367	13	BG983576	BG983576 IL5-CN006
	34	37.2	1.7	318	10	AV644201	AV644201 AV644201
	35	37	1.7	348	12	BF717547	BF717547 fd46g02.y
	36	37	1.7	606	14	BQ806505	BQ806505 WHE3579_H
	37	37	1.7	838	14	BQ671037	BQ671037 AGENCOURT
	38	37	1.7	1016	14	BQ059614	BQ059614 AGENCOURT
	39	36.8	1.7	668	9	AU070770	AU070770 AU070770
	40	36.6	1.7	313	12	BE804382	BE804382 sr79a10.y
	41	36.6	1.7	436	10	AV637236	AV637236 AV637236
	42	36.6	1.7	471	10	AV640279	AV640279 AV640279
	43	36.6	1.7	478	10	AV643425	AV643425 AV643425
	44	36.6	1.7	497	10	BE552689	BE552689 946083H03
	45	36.6	1.7	509	12	BF252639	BF252639 EST419901

ALIGNMENTS

```
RESULT 1
BH388186/c
LOCUS       BH388186                 537 bp    DNA     linear   GSS 11-DEC-2001
DEFINITION  AG-ND-124P3.TF ND-TAM Anopheles gambiae genomic clone AG-ND-124P3,
            DNA sequence.
ACCESSION   BH388186
VERSION     BH388186.1  GI:17334327
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
  TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-124P3.TR
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
```

```
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 For
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-124P3"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      153 a    112 c     94 g    178 t
ORIGIN

  Query Match             5.1%;  Score 110;  DB 17;  Length 537;
  Best Local Similarity   55.9%;  Pred. No. 1.5e-20;
  Matches  209;  Conservative    0;  Mismatches  165;  Indels    0;  Gaps    0;

Qy        527 TCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGG 586
              | ||  ||  |||||         || || |  | |  | | || | | | |  ||||| |
Db        374 TGGTGAAGGATTTCTTCAAAGGAAATGAGTTTTACTACTTTGTATTCCAGGATTTCAAAG 315

Qy        587 ACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACA 646
              | |||||| |||||  | |||| ||      ||| | ||||| |  || || ||||| || |
Db        314 ATGTTCGTCTGGTAGGTACTCCGCCGTCATCTATTGGTAAATATGGTGGTGATACTGATA 255

Qy        647 ACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCG 706
              |||||    |||||  | ||||| || ||||||    || ||||| || ||||| | |
Db        254 ACTGGGAATGGCCAAGACACACAGGAGACTTCTCTGTTTTCCGTGTTTATGCTGATAAAA 195

Qy        707 ACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTG 766
              |    |     || || |||||          ||||      || |  |||||        |
Db        194 ATGGCAATCCTGCTGAATATTCAGCTAACAACGTTCCTTTAAAGCCTAAGCATCATTTAC 135

Qy        767 CCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCA 826
              |  | ||   |   ||||  | ||| |||   || |  ||| ||||      | || |   || || |
Db        134 CAATTTCACTAAAAGGAAATAAGCCTGGTGATTTTGCAATGATTGTAGGATATCCTGGTA 75

Qy        827 GTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATC 886
               ||   |   |  ||||| |||||||  ||   | |||  |       |      || ||  ||
Db         74 CAACAAACAGATACCTTACTTCTTTCGGAATCGAACAAATGGTAAGCAAAGATTACCCGG 15

Qy        887 CTCGTATCGAAGTT 900
              |  |   | |||| |
Db         14 CATGGGTTGAAGCT 1


RESULT 2
BH392840/c
LOCUS       BH392840                 591 bp    DNA     linear   GSS 11-DEC-2001
DEFINITION  AG-ND-162P14.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162P14
            , DNA sequence.
ACCESSION   BH392840
VERSION     BH392840.1  GI:17338981
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
  TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-162P14.TR
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 For
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-162P14"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      170 a    122 c    105 g    194 t
ORIGIN

  Query Match             5.1%;  Score 110;  DB 17;  Length 591;
  Best Local Similarity   55.9%;  Pred. No. 1.6e-20;
  Matches  209;  Conservative    0;  Mismatches  165;  Indels    0;  Gaps    0;

Qy        527 TCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGG 586
              | ||  ||  |||||         || || |  | |  | | || | | | |  ||||| |
Db        374 TGGTGAAGGATTTCTTCAAAGGAAATGAGTTTTACTACTTTGTATTCCAGGATTTCAAAG 315

Qy        587 ACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACA 646
              | |||||| |||||  | |||| ||      ||| | ||||| |  || || ||||| || |
Db        314 ATGTTCGTCTGGTAGGTACTCCGCCGTCATCTATTGGTAAATATGGTGGTGATACTGATA 255

Qy        647 ACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCG 706
              |||||    |||||  | ||||| || ||||||    || ||||| || ||||| | |
Db        254 ACTGGGAATGGCCAAGACACACAGGAGACTTCTCTGTTTTCCGTGTTTATGCTGATAAAA 195

Qy        707 ACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTG 766
              |    |     || || |||||          ||||      || |  |||||        |
Db        194 ATGGCAATCCTGCTGAATATTCAGCTAACAACGTTCCTTTAAAGCCTAAGCATCATTTAC 135

Qy        767 CCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCA 826
              |  | ||   |   ||||  | ||| |||   || |  ||| ||||      | || |   || || |
Db        134 CAATTTCACTAAAAGGAAATAAGCCTGGTGATTTTGCAATGATTGTAGGATATCCTGGTA 75

Qy        827 GTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATC 886
               ||   |   |  ||||| |||||||  ||   | |||  |       |      || ||  ||
Db         74 CAACAAACAGATACCTTACTTCTTTCGGAATCGAACAAATGGTAAGCAAAGATTACCCGG 15

Qy        887 CTCGTATCGAAGTT 900
              |  |   | |||| |
Db         14 CATGGGTTGAAGCT 1


RESULT 3
BH386991/c
LOCUS       BH386991                 500 bp    DNA     linear   GSS 10-DEC-2001
DEFINITION  AG-ND-148L11.TR ND-TAM Anopheles gambiae genomic clone AG-ND-148L11
            , DNA sequence.
ACCESSION   BH386991
VERSION     BH386991.1  GI:17333133
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
```

```
TITLE       Direct Submission of BAC-end sequences from Anopheles gambiae
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-148L11.TF
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..500
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-148L11"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      143 a    105 c     90 g    162 t
ORIGIN

  Query Match             5.1%;  Score 108.2;  DB 17;  Length 500;
  Best Local Similarity   57.6%;  Pred. No. 4.7e-20;
  Matches  194;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy    527 TCGTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGG 586
Db    369 TGGTGAAGGATTTCTTCAAAGGAAATGAGTTTTACTACTTTGTATTCCAGGATTTCAAAG 310

Qy    587 ACGTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACA 646
Db    309 ATGTTCGTCTGGTAGGTACTCCGCCGTCATCTATTGGTAAATATGGTGGTGATACTGATA 250

Qy    647 ACTGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCG 706
Db    249 ACTGGGAATGGCCAAGACACACAGGAGACTTCTCTGTTTTCCGTGTTTATGCTGATAAAA 190

Qy    707 ACAACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTG 766
Db    189 ATGGCAATCCTGCTGAATATTCAGCTAACAACGTTCCTTTAAAGCCTAAGCATCATTTAC 130

Qy    767 CCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCA 826
Db    129 CAATTTCACTAAAAGGAAATAAGCCTGGTGATTTGCAATGATTGTAGGATATCCTGGTA 70

Qy    827 GTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGA 863
Db     69 CAACAAACAGATACCTTACTTCTTTCGGAATCGAACA 33

RESULT 4
BH375201
LOCUS       BH375201              418 bp    DNA     linear   GSS 10-DEC-2001
DEFINITION  AG-ND-173C3.TR ND-TAM Anopheles gambiae genomic clone AG-ND-173C3,
            DNA sequence.
ACCESSION   BH375201
VERSION     BH375201.1  GI:17321343
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
  TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..418
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-173C3"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      145 a     81 c     70 g    122 t
ORIGIN

  Query Match             4.9%;  Score 103.8;  DB 17;  Length 418;
  Best Local Similarity   54.6%;  Pred. No. 8e-19;
  Matches  207;  Conservative    0;  Mismatches  172;  Indels    0;  Gaps    0;

Qy    530 TAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACG 589
Db     30 TAAAGTCCATGTATGATGGTAACAAATATTATGCGTTTATCATTGAGACTTTCAAAGATA 89

Qy    590 TTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACT 649
Db     90 TCCGTTTAGTAGGAGCTCCACCTCAATCTATTGGTAAATTTGGTTCTGATACAGACAACT 149

Qy    650 GGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACA 709
Db    150 GGGTATGGCCAAGACATACAGGAGACTTCTCTATGTTCCGTATCTATGCAGACAAAAACA 209

Qy    710 ACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCG 769
Db    210 ATAAACCTGCAGAATATTCTAAAGATAACATTCCATACACTCCGAAATATTCACTTCCGG 269

Qy    770 TATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTA 829
Db    270 TATCTGTAAAAGACCTTAAAGAAGGTGATTTCACATTCGTTTTCGGATTCCCGGGAAAAA 329

Qy    830 CGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTC 889
Db    330 CAACTGAATACCTTCCTTCAATTGCTGTAGAAAAGATCATCAACGATACAGATCCTGCTA 389

Qy    890 GTATCGAAGTTCGCGGTAT 908
Db    390 AAATTACAGTACGTGATAT 408

RESULT 5
BH400391/c
LOCUS       BH400391              426 bp    DNA     linear   GSS 11-DEC-2001
DEFINITION  AG-ND-147H4.TF ND-TAM Anopheles gambiae genomic clone AG-ND-147H4,
            DNA sequence.
ACCESSION   BH400391
```

```
VERSION     BH400391.1  GI:17346607
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
  TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-147H4.TR
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 For
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..426
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-147H4"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      111 a     79 c     76 g    160 t
ORIGIN

  Query Match             3.4%;  Score 72.4;  DB 17;  Length 426;
  Best Local Similarity   57.8%;  Pred. No. 1.2e-09;
  Matches  149;  Conservative    0;  Mismatches  106;  Indels    3;  Gaps    1;

Qy       1756 GGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAGAAGCAGGATCCTAAG 1815
              |||  |  | | || ||  |||| |||     | ||  |  |  ||||| |  |
Db        258 GGTATCAAGCAGAATTACTATACCACAATGGAAGGTATGATTAAGAAGTACAAGAAAGGT 199

Qy       1816 AGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGCACCAAAAACTATGGT 1875
              ||| || || |    |  |     |     | || || || |     | ||||| ||||||
Db        198 GACGAAGAATTCGATCTTCCACAAGGACTTCTTGATCTTTATAAAAAGAAAAATTATGGT 139

Qy       1876 ---CGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAACAACGACATCACG 1932
                 | | | | |  |||| || || ||| |    ||||| || || ||||| || ||
Db        138 ATGTACAAAGACAAAGACGGGCAACTTCATGTAAACTTCCTTTCTAATAACGATATTACA 79

Qy       1933 GGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATCGGTCTTGCTTTC 1992
              || |||||||| |||    ||  |   |||||     ||||  | || || |||||||| ||
Db         78 GGAGGTAACTCAGGTTCTCCAATTATCGATGGTTACGGAAGACTTATAGGTCTTGCATTT 19

Qy       1993 GATGGCAACTGGGAAGCT 2010
              || || ||| | ||||||
Db         18 GACGGAAACAGTGAAGCT 1


RESULT 6
BH375688/c
LOCUS       BH375688                 697 bp    DNA     linear   GSS 10-DEC-2001
DEFINITION  AG-ND-120J9.TR ND-TAM Anopheles gambiae genomic clone AG-ND-120J9,
            DNA sequence.
ACCESSION   BH375688
VERSION     BH375688.1  GI:17321830
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
  TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-120J9.TF
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..697
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-120J9"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      214 a    133 c    122 g    228 t
ORIGIN

  Query Match             3.1%;  Score 66.4;  DB 17;  Length 697;
  Best Local Similarity   50.3%;  Pred. No. 9.1e-08;
  Matches  197;  Conservative    0;  Mismatches  186;  Indels    9;  Gaps    1;

Qy         16 AAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGGGTAGCCAAAGCCGAC 75
              |||| || ||  | | |   |||  | ||      |   |   | |    | ||  ||   |
Db        475 AAAAACATCCTATTAGCACTAGCTTTACTCCCCGCAGTGACGGCTTTTGCTCAACAGGCG 416

Qy         76 AAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATGCGTGAGCTC 135
                ||| ||||||  | |   |   ||   | |||  |  || |          ||||    ||  |
Db        415 GGAGGAATGTGGATTCCTACAGAGTTAAATGAAAAGGA---------AATGAAGGAATTG 365

Qy        136 GGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATTGCCAATGCC 195
              ||| |  | |  |  |      | |       | | | ||    | || ||    |||     ||||
Db        364 GGCATGAAGATTTCTGCGAAAGATATTTTCAATACTCAAAAACCTAGTATTAAAGATGCT 305

Qy        196 GTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTTT 255
              || ||     |||    || || |||||  | |    |    |     |  ||    | ||   || | ||
Db        304 GTAGTGCAATTCAACGGAGGCTGTACTGCTGAAATTATTTCTCCAAAAGGTTTGTTATTG 245

Qy        256 ACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTAT 315
              || || ||||||||| || | |||      || ||     ||     ||| ||   |   | |||   |
Db        244 ACAAATCACCACTGTGGCTTCGGACAGATTCAAGCTCATTCCACTGTACAAAATGACCTT 185

Qy        316 CTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCGATTCCGGGTCTTTCC 375
              ||   |   ||||||||      |       |    ||||| || || ||||| |   || ||| |
Db        184 CTAAGTAATGGTTTCTGGGCAAAAAACATGGGAGAAGAACTTCCAAACCCTGGTGTGGTT 125

Qy        376 GTGAAGTATCTGCGCAAGATCGTGAAGGTAAC 407
              || |  |    |  | | ||       |||| ||
Db        124 GTAGATTTCATCACCGATATTAAAGAGGTGAC 93
```

```
RESULT 7
AW774807
LOCUS       AW774807                398 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  EST333958 KV3 Medicago truncatula cDNA clone pKV3-24C3, mRNA
            sequence.
ACCESSION   AW774807
VERSION     AW774807.1  GI:7718724
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 398)
  AUTHORS   VandenBosch,K., Hurt,J., Moore,J., Beremand,P., Peng,H., Ellis,L.,
            Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
            Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula after Rhizobium inoculation
  JOURNAL   Unpublished (1999)
COMMENT     Contact: VandenBosch K
            Department of Plant Biology
            University of Minnesota
            220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
            Tel: 612 624 2755
            Fax: 612 625 1738
            Email: kvandenb@cbs.umn.edu
            Texas A&M EST name:T258226e
            TIGR sequence name:MTEBG14TK
            More information is available at.
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1..398
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pKV3-24C3"
                     /clone_lib="KV3"
                     /tissue_type="Seedling roots"
                     /dev_stage="3 days post-inoculation with Sinorhizobium
                     meliloti"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts.  Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
BASE COUNT      137 a     64 c     81 g    116 t
ORIGIN

  Query Match             2.6%;  Score 55;  DB 10;  Length 398;
  Best Local Similarity   50.6%;  Pred. No. 0.00013;
  Matches  133;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy         67 AAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGATCGAATG 126
              | ||||||  |||| |||||| |||| |         | | |        ||||        |||
Db         28 AGAGCCGATGAAGGAATGTGGTTCCTGATGTTTATCGAAAGATTAAATCACAGGGATATG 87

Qy        127 CGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCGTCCATT 186
              |   |  | ||||      |  |     ||       | |||||  |  ||||  |      | |
Db         88 CAAAAAATGGGCTTGCAATTGACAGCCGAAGAAATTTACAGCATTAACAATCACAGCTTA 147

Qy        187 GCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGC 246
                  ||||  |  ||      |||   ||| || ||||| |        || ||   |     ||
Db        148 AAAGATGCGATTGTACAATTCAATGGGGGTTGTACTGCAGAAATCGTTTCAAACAGCGGT 207

Qy        247 CTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGAT 306
               || |  |||||||  |||||||| || ||   | || ||          |       || |  ||
Db        208 TTGGTTCTTACCAATCACCACTGTGGTTATAGCGCGATTGCAGAACTTTCAACTGCAGAA 267

Qy        307 CACGACTATCTGCGCGATGGTTT 329
              ||  |  || |     |||||||
Db        268 CAAAATCATTTAAAAAATGGTTT 290

RESULT 8
BH164253/c
LOCUS       BH164253               1005 bp    DNA     linear   GSS 24-SEP-2001
DEFINITION  ENTTG63TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH164253
VERSION     BH164253.1  GI:15737691
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 1005)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 15
            High quality sequence stop: 487.
FEATURES             Location/Qualifiers
     source          1..1005
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      186 a    314 c    336 g    169 t
ORIGIN

  Query Match             2.3%;  Score 49.2;  DB 17;  Length 1005;
  Best Local Similarity   56.0%;  Pred. No. 0.012;
  Matches   93;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        195 CGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGGCCTGATCTT 254
              ||  ||   |    ||  || ||  ||||||||        ||||||   | |||||| |  |
Db        229 CGCCGTGGTGCGCGCCGGTGGTGGTACCGGTTCGTTCGTGTCCGCCGACGGCCTGCTGCT 170

Qy        255 TACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACGGTGGATCACGACTA 314
               |||||||||||     |         || |   ||||||  | |  ||| ||     || ||||||
Db        169 GACCAACCACCACGTGGCGATGGGCGTGATCCAGTACAACAGCTCGCCCGAGCACGACCT 110

Qy        315 TCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360
                |    | |||  |||  |    |     |           |   || ||||  |||
Db        109 GATCACTGGTGGCTTCATCGCCAACGGCCGTGCCGACGAGCGCCCG 64
```

```
RESULT 9
BH400866/c
LOCUS       BH400866                 555 bp    DNA     linear   GSS 11-DEC-2001
DEFINITION  AG-ND-125M4.TR ND-TAM Anopheles gambiae genomic clone AG-ND-125M4,
            DNA sequence.
ACCESSION   BH400866
VERSION     BH400866.1  GI:17347082
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
  TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-125M4.TF
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..555
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-125M4"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      152 a    105 c     94 g    204 t
ORIGIN

  Query Match             2.0%;  Score 42.8;  DB 17;  Length 555;
  Best Local Similarity   66.0%;  Pred. No. 0.6;
  Matches   62;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        926 AGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGA 985
               | |||    ||| |   | ||| | || |||||||| ||||||||      || |||||
Db        547 ATGAAAATATGCGTACTGATAATGCAACCCGTATTAAATATGCATCTAAATATGCATCCG 488

Qy        986 GTGCTAACTATTGGAAGAATTCGATCGGTATGAA 1019
                |||||||||||| | ||    ||| |||| | |
Db        487 TGGCTAACTATTGGGAAAAAATGATGGGTAAGTA 454


RESULT 10
AI405287
LOCUS       AI405287                 692 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  GH25242.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH25242 5prime, mRNA sequence.
ACCESSION   AI405287
VERSION     AI405287.1  GI:4248374
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 252  row: D  column: 6
            High quality sequence stop: 496.
FEATURES             Location/Qualifiers
     source          1..692
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH25242"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      147 a    210 c    181 g    153 t      1 others
ORIGIN

  Query Match             2.0%;  Score 42.2;  DB 9;  Length 692;
  Best Local Similarity   49.3%;  Pred. No. 1;
  Matches  110;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy       1470 GGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGC 1529
              ||   ||   |||  |  | |    ||||   |  ||||||| |||  ||     |||||
Db        254 GGCCGCTATCAGCCGCCTGCTTGCTGCCTCCCAGAAGTCCAGCGACTCGGTTCTGTTTGT 313

Qy       1530 CAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCG 1589
              | || |  |   ||||| ||    ||||| |  || | ||||| |  |   |     ||| |||
Db        314 CCAGACCACGGAGAAGGACGTGGCAGCATCGCGTGCCAAGCGTGATACCATTGCCGCTTC 373

Qy       1590 CGCTATTCAGGCCGATGCGATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTT 1649
              | | |    |  || |       |||||     ||| | || || |       |    |
Db        374 CACCACCAATCCCTACAACCTGGCCGTCTACTACGGCAGCGACTACCCGGTGATCTTCAA 433

Qy       1650 CTTTGCCGGTTTGCGTGAGATGTACCCCGGACGTGCTCTGCCG 1692
              | |   |   | |      | |   |  |||||   |||||| |
Db        434 CATCATCCTGTGGTTCATGGTCGTCTTCGGACTGTCTCTGCTG 476


RESULT 11
BH391233
LOCUS       BH391233                 721 bp    DNA     linear   GSS 11-DEC-2001
DEFINITION  AG-ND-141L2.TF ND-TAM Anopheles gambiae genomic clone AG-ND-141L2,
            DNA sequence.
ACCESSION   BH391233
VERSION     BH391233.1  GI:17337374
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
  TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-141L2.TR
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
```

```
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 For
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..721
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-141L2"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      269 a    122 c    147 g    183 t
ORIGIN

  Query Match             1.9%;  Score 41.2;  DB 17;  Length 721;
  Best Local Similarity   47.3%;  Pred. No. 2.1;
  Matches  124;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy        924 GAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAGCAAGTATGCTCA 983
              ||||  |      |||        |||  | |||||   |  |      || ||      || |||||
Db         28 GAAGAAGCATATGGATAAAGACGATGCTACAAGATTAGCTTACGCGTCTAACTATGCAAG 87

Qy        984 GAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGT 1043
                 |||||||||||||| ||       ||| |   |      |                 |
Db         88 TGTTGCTAACTATTGGAAAAACAGAGCCGGAACAATTGAAGCGGGTTACAAAAACGTATC 147

Qy       1044 GATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCAA 1103
               || |||    ||      |    |   || |  |   || |   |  | |||       |||     |  ||
Db        148 AATTGGTGACAAAAAAGAAGTTGAGAAAAAATACCAACAATGGGCTGATAAAGCTGAAAA 207

Qy       1104 GAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAA 1163
               |  ||||| |||||  |||||||   |     |     ||          || |||  ||      |  ||
Db        208 CAAAGCTGTTTATGGAAATGTATTAGCAAATACTGATGCATACTACAAGCAAATTTCTAA 267

Qy       1164 GGCCAACCGTGAGATGACTTAT 1185
                  ||     ||||  |  ||||
Db        268 CAGAAATATCGAGAAAAAATTAT 289


RESULT 12
BH371846
LOCUS       BH371846               822 bp    DNA     linear   GSS 10-DEC-2001
DEFINITION  AG-ND-162M17.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162M17
            , DNA sequence.
ACCESSION   BH371846
VERSION     BH371846.1  GI:17317971
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 822)
  AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
  TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-162M17.TR
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 For
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..822
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-162M17"
                     /clone_lib="ND-TAM"
                     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      235 a    142 c    155 g    290 t
ORIGIN

  Query Match             1.8%;  Score 39.4;  DB 17;  Length 822;
  Best Local Similarity   65.2%;  Pred. No. 7.6;
  Matches   58;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       2007 AGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACAT 2066
              ||||  |  ||  ||||||||  |    ||  |||||  |   |  ||    |  |  ||  |  |||  |||
Db          1 AGCTTTAAGCGGTGACATTGTTTTTGAACCTAAATTACAAAGAACGATTAACGTAGACGT 60

Qy       2067 CCGCTACGTTCTCTTCATGATTGACAAAT 2095
                | |||||  ||  |      |  ||  |||||  |
Db         61 TAGATACGTACTTTGGGTAATCGACAAGT 89


RESULT 13
AL728310
LOCUS       AL728310               606 bp    mRNA    linear   EST 18-APR-2002
DEFINITION  AL728310 Danio rerio embryonic inner ear subtracted cDNA Danio
            rerio cDNA clone BN0AA098ZA08 3', mRNA sequence.
ACCESSION   AL728310
VERSION     AL728310.1  GI:20192914
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Coimbra,R., Weil,D., Brottier,P., Blanchard,S., Levi,M., Hardelin
            ,J.P., Weissenbach,J. and Petit,C.
  TITLE     A subtracted cDNA library from the zebrafish (Danio rerio)
            embryonic inner ear
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1..606
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="BN0AA098ZA08"
                     /clone_lib="Danio rerio embryonic inner ear subtracted
                     cDNA"
                     /tissue_type="inner ear"
                     /dev_stage="embryonic"
                     /note="subtracted cDNA library"
BASE COUNT      180 a    100 c    157 g    169 t
ORIGIN

  Query Match             1.8%;  Score 39;  DB 9;  Length 606;
  Best Local Similarity   47.7%;  Pred. No. 8.2;
```

```
  Matches  114;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy      852 GGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTTCGCGGTATCAA 911
             ||  | || ||||  ||     |      | |||  ||||| |||       |    | |||
Db      116 GGAAGAGGCAGATTATAAGAGACTACATAACTTTCCTCTTATAAGGCATACTGATATGCC 175

Qy      912 GCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAATATGCCAG 971
              |||  ||  ||  |||  |||||          |     |   |   ||    || | |||||
Db      176 TGAAGAGATGCGGGTGGAGACCATGGAGCTCTGTGTTACAGCGTGTGAAAAGTTTGCCAC 235

Qy      972 CAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAACCGCGGTCTCGC 1031
              ||  |||        |||||             ||| |  ||||| | || |||      ||    | |
Db      236 TAATAATGAGAGCGCAGCTAAAATGATAAAGGAGTCGATGGATAAGAAGTTTGGGAGCTC 295

Qy     1032 TCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCAGACTGGATCC 1090
             | | |  ||  ||||| ||             ||        ||||    ||    | || |  || ||
Db      296 TTGGCATGTGGTGATTGGCGAAGGCTTTGGGTTCGAAATCACACATGAAGTCAAGAACC 354


RESULT 14
BH448937/c
LOCUS       BH448937                 687 bp    DNA     linear   GSS 12-DEC-2001
DEFINITION  BOGZL67TR BOGZ Brassica oleracea genomic clone BOGZL67, .DNA
            sequence.
ACCESSION   BH448937
VERSION     BH448937.1  GI:17634648
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGZL67TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGZL67"
                     /clone_lib="BOGZ"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      209 a    117 c    159 g    202 t
ORIGIN

  Query Match             1.8%;  Score 38.6;  DB 17;  Length 687;
  Best Local Similarity   53.7%;  Pred. No. 12;
  Matches   80;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy     1457 TCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGACA 1516
            ||||||     |||||   |   |  ||||| ||||      ||| | || |||    |   ||
Db      685 TCGACACTTTTGTGGGATCAAAAGACGACCAGTTGAGATCCAAGATCTAGTTAGTTAACC 626

Qy     1517 AGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGTAA 1576
             ||  |||  |||||| ||||    |      |||    | ||| |||||  |    | | | |
Db      625 AGTGAAACATTGCCATCGCTAATCACTGTGATTAAGGAGTTGAGCTCTAATATATCATGC 566

Qy     1577 TAGCTGCTGCTCGCGCTATTCAGGCCGAT 1605
               ||  ||| | |    | |  |||    ||
Db      565 ATGCAACTGTTAGGCATCTATAGGATTAT 537
```

```
RESULT 15
AW375157
LOCUS       AW375157                 184 bp    mRNA    linear   EST 04-FEB-2000
DEFINITION  QV2-CT0121-280999-001-e01 CT0121 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW375157
VERSION     AW375157.1  GI:6879811
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV2&t2=QV2-CT0121-
            280999-001-e01&t3=1999-09-28&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 18
            High quality sequence stop: 88.
FEATURES             Location/Qualifiers
     source          1. .184
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0121"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       29 a     53 c     58 g     44 t
ORIGIN

  Query Match             1.8%;  Score 38.4;  DB 10;  Length 184;
  Best Local Similarity   51.1%;  Pred. No. 5.7;
  Matches   90;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy      612 CAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCGCGTCACACGGG 671
            ||||| |||   | || |     ||    |   |||      || || ||||  |    |  |
Db        8 CAGCTTTGTCCTTCAGCTGCTTGGGCTTCTGGAACTTGGGGTGGGGCCCTGCTCCTGGAT 67

Qy      672 CGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCCGAATACAGCAA 731
            | |||||  |        | | |||  |   | |    |  |||| ||    ||    | |||| |
Db       68 CAACTTCTCCACGAAGCTCTTGTCGGTGGCTTTGGGAACCAGCACTCCTCGTCCAGCCA 127

Qy      732 GGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAAGGCTACA 787
            | ||  |  |||     ||| |  |  ||||  |          || ||||| |||| ||
Db      128 GGCCAGAATGCCCGGGGGGCCTGCTGGCTTCTCAATGAGGTCGATGCAGGGCTGCA 183


Search completed: May 23, 2003, 05:18:34
Job time : 2069 secs
```

P
D


```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         May 23, 2003, 08:43:01 ; Search time 3929.8 Seconds
                                           (without alignments)
                                           5272.840 Million cell updates/sec

Title:          US-10-008-355-2
Perfect score:  3719
Sequence:       1 MQMKLKSILLGAALLLGASG..........LFMIDKWGQCPRLIQELKLI 712

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters:      4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110400_2535/app_query.fasta_1.1230
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_2983_@runat_16052003_110400_2535 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	907.5	24.4	10689	1	AE004008	AE004008 Xylella f
c	2	186.5	5.0	253217	2	AC016590	AC016590 Homo sapi
	3	175	4.7	10811	1	AE011732	AE011732 Xanthomon
	4	147.5	4.0	50372	1	AE014150	AE014150 Streptoco
	5	146.5	3.9	10889	1	AE006549	AE006549 Streptoco
	6	144.5	3.9	15569	1	AE000733	AE000733 Aquifex a
c	7	137.5	3.7	162754	8	OSJN00169	AL662969 Oryza sat
	8	133.5	3.6	4004	1	MSP207AAM	X55799 M.sp 207 al
	9	133	3.6	10292	1	AE008428	AE008428 Streptoco
c	10	133	3.6	11362	2	SPNEU1923	AL449945 Streptoco
	11	131.5	3.5	14169	1	AE009858	AE009858 Pyrobacul
	12	130	3.5	2652	6	A81135	A81135 Sequence 1
	13	130	3.5	2652	6	AR065782	AR065782 Sequence
	14	130	3.5	2652	6	E26074	E26074 Novel valS.
	15	130	3.5	9827	1	AE013500	AE013500 Methanosa
c	16	130	3.5	10029	1	AE010493	AE010493 Fusobacte
	17	129	3.5	4141	1	CLOCOLA	D13791 C. perfring
c	18	129	3.5	6004	6	BD003707	BD003707 Polynucle
	19	129	3.5	11280	1	AE007366	AE007366 Streptoco
c	20	129	3.5	26703	3	CBRG45O11	AC084652 Caenorhab
c	21	128.5	3.5	291150	1	AP003135	AP003135 Staphyloc
c	22	128.5	3.5	342600	1	AP003363	AP003363 Staphyloc
	23	128	3.4	11264	1	AE009964	AE009964 Streptoco
	24	127	3.4	193188	2	AC006884	AC006884 Caenorhab
	25	126.5	3.4	90348	1	AF497482	AF497482 Micromono
	26	126	3.4	3084	1	LAU62096	U62096 Lactobacill
c	27	126	3.4	11925	1	AE013596	AE013596 Methanosa
	28	125.5	3.4	52900	1	AE014138	AE014138 Streptoco
	29	124.5	3.3	9389	1	AE008276	AE008276 Agrobacte
c	30	124.5	3.3	11199	1	AE009339	AE009339 Agrobacte
	31	124	3.3	36380	3	AF106581	AF106581 Caenorhab
	32	123.5	3.3	2142	1	AF039313	AF039313 Moraxella
	33	123.5	3.3	10029	1	AE010835	AE010835 Methanosa
c	34	123.5	3.3	10296	1	U67538	U67538 Methanococc
c	35	123	3.3	10181	1	AE001761	AE001761 Thermotog
	36	123	3.3	340806	1	NMA1Z2491	AL162752 Neisseria
	37	122.5	3.3	4716	6	AX415417	AX415417 Sequence
	38	122.5	3.3	4731	6	AX413695	AX413695 Sequence
	39	122.5	3.3	7009	14	AF296095	AF296095 Porcine t
	40	122.5	3.3	7858	1	U35629	U35629 Lactococcus
c	41	122.5	3.3	10826	1	SPU49397	U49397 Streptococc
c	42	122.5	3.3	239050	1	AL596169	AL596169 Listeria
	43	122.5	3.3	299850	1	AP001514	AP001514 Bacillus
	44	122.5	3.3	349980	6	AX413017	AX413017 Sequence
c	45	122.5	3.3	349980	6	AX417045	AX417045 Sequence

ALIGNMENTS

RESULT 1

```
AE004008/c
LOCUS       AE004008               10689 bp    DNA     linear   BCT 15-JUN-2001
DEFINITION  Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.
ACCESSION   AE004008 AE003849
VERSION     AE004008.1  GI:9106961
KEYWORDS    .
SOURCE      Xylella fastidiosa 9a5c.
  ORGANISM  Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 10689)
  AUTHORS   Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
   PUBMED   10910347
REFERENCE   2  (bases 1 to 10689)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
            Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
            13083-970, Brazil
FEATURES             Location/Qualifiers
     source          1. .10689
                     /organism="Xylella fastidiosa 9a5c"
                     /db_xref="taxon:160492"
                     /clone="9a5c"
     gene            complement(119. .367)
                     /gene="XF1871"
     CDS             complement(119. .367)
                     /gene="XF1871"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84677.1"
                     /db_xref="GI:9106962"
                     /translation="MTMTLKEDKTEIIDEQRIQIAKLLSETAKIQAEIHEVNAHTQKL
                     MKETLKLTTESKWYPVVVGSSLIAAGATAATLFIKLFH"
     gene            complement(430. .618)
                     /gene="XF1872"
     CDS             complement(430. .618)
                     /gene="XF1872"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84678.1"
                     /db_xref="GI:9106963"
                     /translation="MMEEARATAPQGGLAASDPQRITDAPKKTGLRKRKYPTTGSASI
                     IVLRHVLTQRQQKRKASP"
     gene            941. .1288
                     /gene="XF1873"
     CDS             941. .1288
                     /gene="XF1873"
                     /note="similar to GI|479340 (percent identity: 36 %/query
                     alignment coverage:  87.8 %/subject alignment coverage:
                     90.2 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF84679.1"
                     /db_xref="GI:9106964"
                     /translation="MSKNKIPNQILTGNWAGFQFINGKLVTPERRTIEEWQLRWLSLT
                     CTLAREWQKMMEEARATAPQGGLAASDPQRITDALQKTVSHKRRHAANETASIIQLRQ
                     VLIQRQQKRKASP"
     gene            1351. .1662
                     /gene="XF1874"
     CDS             1351. .1662
                     /gene="XF1874"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84680.1"
                     /db_xref="GI:9106965"
                     /translation="MTMTLKEDKTEIIDEQKIQIAKLISETAKIQAEIHEVTTHTQEM
                     NAHINEMNARTQEMNARTQKLIEETQKVSKESQWYPVVVASGLLAAGATAATLFIKLF
                     H"
     gene            complement(1839. .2918)
                     /gene="XF1875"
     CDS             complement(1839. .2918)
                     /gene="XF1875"
                     /note="similar to SP|P03626 (percent identity: 36 %/query
                     alignment coverage:  88.0 %/subject alignment coverage:
                     105.0 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="phage-related protein"
                     /protein_id="AAF84681.1"
                     /db_xref="GI:9106966"
                     /translation="MLNLVTGAPGNGKTLYAVDWLIRQIEIDKSLVKSGAVPRSFYTD
                     IEGFDVEAVRRLTGYVVQSAPEDWRTTPQGSVIVYDEAHRMFPAGRPGRSDDPRVCDL
                     DTHRHGGYDLMFVTQWPTKIHHELRRLVGEHVHLNRAMGLQTAGLYRWSRAQDDPYDI
                     HQREKAEEEVWKFPKDRYALYASSTLHTVSHKFRIPKKVWSALSVCVTCSVIGLVFWH
                     YYSPAHLSEAASSVAGASGQASLRAAPASLSSSRSLVSGMRTYAVLETESAPTLSGCV
                     SSDRSCRCFNTDGYQIDMSVVECRRLLASPLPFNVYHAYVTSSSSAAAVSSSPSSSPS
                     SSSVSLSSSVSSAPILGSSSVPSSH"
     gene            complement(2918. .3202)
                     /gene="XF1876"
     CDS             complement(2918. .3202)
                     /gene="XF1876"
                     /note="similar to SP|P03625 (percent identity: 36 %/query
                     alignment coverage:  85.1 %/subject alignment coverage:
```

```
                     86.0 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="phage-related protein"
                     /protein_id="AAF84682.1"
                     /db_xref="GI:9106967"
                     /translation="MFSWLAMLLRNILGQTVARGLVGAGLALVTTVPLIPLVTSALNL
                     IVSKMSGISADVLNIALLMGFGEALSIIGSAMLTRLALQSLHVGIVKATT"
     gene            complement(3195. .4412)
                     /gene="XF1877"
     CDS             complement(3195. .4412)
                     /gene="XF1877"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84683.1"
                     /db_xref="GI:9106968"
                     /translation="MCSRSFFLYFISALGLTLVISFPFCLFAQNVPPTVEVVPTSVRY
                     SSVISDGVSTSATFEARSTALVNGVRYYTVPVDISASTLGSLAKVAVRRGMAFYNVYS
                     MLSGLINGAGWVIDELTHEVMSGPALKEIPVGTVAWWWQRPGDGHVFYSVTPQGLIAP
                     INAYQSSLSPPQPLVTSSGPSSVSSERWLYHLEGGGEYIGHLSKTDQSVPDYSSGLPP
                     SVILDTDLGQLVRSDPSTVNAVLTDSQTGAVLLTPEIVSALNKLRRSLEDELKASHAP
                     DQQPSSGGASSPPSSSGTAWPSFCSWASVVCDFIDWVKSDEFLKKPLVPPDVPYVDKL
                     PQAKTWSSGLGEGACPSPTTIPIEFSGYKTSVEISYQPFCDFAALMRPVVIVIATILA
                     AYIAGGFRGVKNV"
     gene            complement(4481. .4702)
                     /gene="XF1878"
     CDS             complement(4481. .4702)
                     /gene="XF1878"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84684.1"
                     /db_xref="GI:9106969"
                     /translation="MKSYIDRGIAFFSLLSTSSLVFAADAAGASFDAGAAVTALGAIS
                     GAVALIGAAKLAPAAISVGWKWLKGAIFG"
     gene            complement(4724. .4924)
                     /gene="XF1879"
     CDS             complement(4724. .4924)
                     /gene="XF1879"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84685.1"
                     /db_xref="GI:9106970"
                     /translation="MAQQLMSLYCTQYDVEARTCSQQAWMVPPSLLPPISYEDVRILL
                     PHIVMCFLVAWGFHFLFTVVRD"
     gene            complement(4931. .5236)

Alignment Scores:
Pred. No.:              3.58e-53       Length:          10689
Score:                  907.50         Matches:         239
Percent Similarity:     49.21%         Conservative:    135
Best Local Similarity:  31.45%         Mismatches:      291
Query Match:            24.40%         Indels:          95
DB:                     1              Gaps:            21

US-10-008-355-2 (1-712) x AE004008 (1-10689)

Qy      1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20
           |||:::      |||   |||:::   :::  |||  |||:::         |||
Db  10499 ATGCGTTTTAACCTGTTGTCGCTTTCCGTTTTAGCTACATTGATAACTGTCGACTCC--- 10443

Qy     21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAsp 40
                |||   :::|||||||||:::       :::|||      |||
Db  10442 ---ACTCATGCGGGCGAAGGCATGTGGGTTCCACAACAGCTT---CCGGAAATTGCTGGC 10389

Qy     41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60
             :::::::::   |||      |||      :::   |||   :::
Db  10388 CCTCTAAAGCAGGCTGGCCTACAACTGTCACCAGAACAATTGTCAAATTTGACAGGTGAC 10329

Qy     61 SerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80
             :::         |||   :::      |||   ||||||            ||||||
Db  10328 CCCATGGGGGCAGTGGTCTCTCTC------GGCAATTGCACCGCGAGTCTGGTTTCACCA 10275

Qy     81 GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100
          :::||||||:::   |||||||||||||||   |||||||||||||||         ||||||
Db  10274 GAAGGACTGGTGATCACTAACCATCATTGCGCTTATGGTGCGATTCAGTTGAACTCTACG 10215

Qy    101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuPro 120
                :::   :::::::::||||||   :::   |||         :::|||:::
Db  10214 CCTAAAAAGAACTTAATCAAAGAAGGGTTCAATGCGTTGACTCAGGCTGATGAAGTCAGT 10155

Qy    121 Ile---ProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysVal 139
                |||         :::      |||   :::|||         ||||||   :::
Db  10154 GCGGGTCCGAATGCGCGTATTTACGTGCTTGAGCAGATTACCGATGTCACCGCTCAAGCT 10095

Qy    140 GluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluVal 159
          :::         |||         ::::::      :::         |||   |||
Db  10094 AAAGCTGCTCTGGCTGCTGCCGGTAATGACCCCTTCAAGCGTACGACGGCATTGGAGACA 10035

Qy    160 Cys------GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleVal 177
                   ||||||:::||||||   |||      ::::::         |||
Db  10034 TTCAGCAAACAAGAAATCGCCAAATGCGAGGAA---GAGCAGGGTTATCGCTGCCAATTC 9978

Qy    178 GluProPheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspVal 197
             |||         |||   |||   :::               |||||||||
Db   9977 TTCAGTTTCGCTGGTGGAAATACCTACAGAGTGTTCAAGAACCTGGAGATCAAGGACGTG 9918

Qy    198 ArgMetValPheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrp 217
          |||:::|||:::||||||      ||||||||||||||||||||||||   |||||||||
Db   9917 CGTTTGGTTTACGCACCGCAGGGGAGTGTGGGTAAGTTCGGGGGCGACGTCGATAACTGG 9858

Qy    218 MetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsn 237
          ||||||||||||||||||||||||||||||   :::|||   |||   |||   |||
Db   9857 ATGTGGCCGCGTCATACCGGTGACTTTTCTTTCTACCGTGCCTACGTCGGTAAAGACGGT 9798

Qy    238 ArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaVal 257
          :::||||||   :::||||||:::|||   ||||||:::|||   ::::::
Db   9797 AAACCAGCGAGCTTCTCTAAAGAAAACATACCCTACCGTCCCAAACACTGGCTGAAATTT 9738

Qy    258 SerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThr 277
          |||   |||            |||:::   |||         |||:::||||||   |||
Db   9737 AGCGATCAACCATTAGGTGATGGTGACTTCGTGATGGTCGCTGGTTACCCAGGACGTACC 9678

Qy    278 AspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArg 297
          :::||||||                           |||   ||||||
Db   9677 AATCGTTATGCTTTGGTG-----------------GCCGAATTTGAGAATACTGCTCAC 9636

Qy    298 IleGluValArgGlyIleLysGln-----------------GlyIleTrpLysGluAla 311
                      |||   |||                     :::   :::   |||
Db   9635 TGGACTTATCCGGTGATTGGCCAGCATTTCAAGAATCTCATTGCTTTGATTGAGGCGGCT 9576

Qy    312 MetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsn 331
                   :::      ::::::||||||||||||      |||         |||
Db   9575 AGCAAACAGAATCCAGACATTCAAGTGAAATATGCGAGCACTTTAGCTGGATTGAATAAC 9516

Qy    332 TyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArg 351
                ||||||      |||         |||   |||::::::   ||||||:::
Db   9515 ACATCGAAAAACTTTGATGGTCAATTGGATGGCTTCCGACGCATTAATGCTATAGGTCAG 9456

Qy    352 LysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGly------------ 367
          |||::::::||||||   |||      |||:::::::::   |||
Db   9455 AAACAGAGTGAAGAAACAGCAGTATTAGCTTGGTTGAAACAGCAGGGCATACGTGGACAC 9396

Qy    368 LysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAla 387
          ::::::                  |||      |||   :::   |||
Db   9395 GAAGCATTGGCTGCGCATCAAACATTGGTTGATCTGACAGAACAATAC------------ 9348
```

```
Qy    388 LysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVal 407
          |||||||||:::::      ::::::              |||  |||:::  |||:::
Db   9347 AAAGCAAATCAGGATCGTGATTTTGTATTGGGTCAG---TTCAACGGAAGTGGTGTCATT 9291

Qy    408 ArgPheAlaGlnPheAlaAsnAlaLeuAla---------ThrAsnProAspAlaHis--- 423
             |||             ||||||          |||    ||||||
Db   9290 GGTGTAGCGGTGAACCTATATCGCTTGGCAATCGAGCGCACAAAATCTGATGCCCAACGT 9231

Qy    424 ------------------------------------AlaGlyIleLeuLysSerLeuAsp 431
                                              |||  ||||||  ::::::
Db   9230 GAAGCGGGATATCAGGAGCGCGATCTGCCAACCATTGAAGGCAACTTAAAGCAGATGGAG 9171

Qy    432 AspLysTyrLysAspTyrLeuProSerLeuAspArgLysVal------------------ 445
                :::   |||||||||   :::||||||::::::
Db   9170 ---------CGACGTTACTTACCGGAAATGGATCGCCAGATGCAGCAATACTGGCTCACC 9120

Qy    446 ------------LeuProAlaMetLeuAspIleValArgArgArgIleProAla------ 459
                      ||||||              |||::::::|||:::   |||
Db   9119 GAATACAACAAGCTTCCT---------------GTCAAACAGCGTGTGGCTGCAATTGAT 9075

Qy    460 -----------AspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
                     |||   :::|||          |||   :::   |||  |||
Db   9074 GTTTGGTTGGGAGATGGCATCCCAGCAACTCTCAAGCGCTTGGGTGACACCAAATTAAGC 9015

Qy    476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
                ::::::
Db   9014 AGTTCTGAGGAAAGG--------------------------------------------- 9000

Qy    496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
                           ||||||     :::  ::::::  |||              :::
Db   8999 ---------------CTGAAGTGGTTTAATGCCGACCGTGCAGCCTTCGAAAGCAGTCAA 8955

Qy    516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAsp 535
          |||||||||:::      :::  :::::::::  |||          |||:::
Db   8954 GATCCAGCGATCCGTTATGCTGTGGCCATCATGCCAGCATTACTGGAGATTGAACGTCAG 8895

Qy    536 AlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArg 555
                           :::  |||  :::   ::::::              |||
Db   8894 AATAAAATCCGTACTGGAGAACTACTTAAAGCGCGTCCGATCTACTTGCAAGCCTTAGCT 8835

Qy    556 GluMetTyrPro------GlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMet 573
          :::                 |||:::  :::     |||||||||  ::::::|||:::
Db   8834 GACTACAACAAGAGTCATGGCAAATTTGTTTACCCAGACGCCAACTCATCATTACGTATT 8775

Qy    574 SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThr 593
          ::::::|||   :::|||||||||   |||:::||||||         |||        |||
Db   8774 ACCTTCGGCCACGTGAAAGGCTATTCACCCAAAGATGGCGTCGAGTACACACCGTTTACG 8715

Qy    594 ThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGlu 613
          |||   :::||||||:::  |||   :::          :::  |||          :::
Db   8714 ACACTGCAGGGTGTCATGGCGAAG---AATACAGGCGTCGAGCCATTCGATTCACCGAAG 8658

Qy    614 AsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsn------Gly 631
          ::::::::::::      :::   |||:::|||      |||:::             |||
Db   8657 TCACTGATTAATGCGATAAAGGCAAAGAGCTACGCCAATTTAGCCGATCAGCGTATTGGC 8598

Qy    632 GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerPro 651
             :::   :::   |||||||||:::   |||||||||||||||||||||||||||
Db   8597 ACTGTCCCTGTGAATTTTCTATCTGACTTGGACATTACTGGCGGCAACTCCGGATCGCCA 8538

Qy    652 ValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMet 671
          |||   |||   :::|||:::|||:::|||||||||||||||||||||||||||::::::
Db   8537 GTCTTAGATGCCCACGGAAAATTGGTGGGTTTAGCATTTGACGGTAATTGGGAGTCTGTC 8478

Qy    672 SerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerValAspIleArgTyr 691
          |||   :::     |||:::|||   :::   |||||||||:::||||||   ||||||
Db   8477 AGTAGTAATTGGGTGTTTGATCCTGTGATGACTCGAACAATTGCAGTTGATAGTCGTTAT 8418

Qy    692 ValLeuPheMetIleAspLysTrpGlyGlnCysProArgLeuIleGlnGluLeuLysLeu 711
          |||   :::::::::  :::                |||   |||:::::||||||   |||
Db   8417 GTGCAGTGGATCATGACTGAAGTTGCACCAGCACCACATCTACTAAAGGAATTGAATTTG 8358
```

RESULT 2

```
AC016590/c
LOCUS       AC016590              253217 bp    DNA     linear   HTG 20-APR-2001
DEFINITION  Homo sapiens chromosome 19 clone CTD-3220F14, WORKING DRAFT
            SEQUENCE, 39 unordered pieces.
ACCESSION   AC016590
VERSION     AC016590.6  GI:13699590
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 253217)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 253217)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Apr 20, 2001 this sequence version replaced gi:7711568.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 940643, BC905667
            Center clone name: CITB-E1_3220F14
            -------
            Summary Statistics
            Consensus quality: 207648 bases at least Q40
            Consensus quality: 225569 bases at least Q30
            Consensus quality: 232428 bases at least Q20
            Estimated insert size: 250510; agarose-fp estimation
            Estimated insert size: 249417; sum-of-contigs estimation
            Quality coverage:  9.03 in Q20 bases; agarose-fp estimation
            Quality coverage:  9.07 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 39 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1019: contig of 1019 bp in length
            *     1020     1119: gap of unknown length
            *     1120     2152: contig of 1033 bp in length
            *     2153     2252: gap of unknown length
            *     2253     3701: contig of 1449 bp in length
            *     3702     3801: gap of unknown length
            *     3802     5053: contig of 1252 bp in length
            *     5054     5153: gap of unknown length
            *     5154     6184: contig of 1031 bp in length
            *     6185     6284: gap of unknown length
            *     6285     7809: contig of 1525 bp in length
            *     7810     7909: gap of unknown length
            *     7910     8986: contig of 1077 bp in length
            *     8987     9086: gap of unknown length
            *     9087    10090: contig of 1004 bp in length
            *    10091    10190: gap of unknown length
            *    10191    11225: contig of 1035 bp in length
            *    11226    11325: gap of unknown length
            *    11326    12365: contig of 1040 bp in length
            *    12366    12465: gap of unknown length
            *    12466    14535: contig of 2070 bp in length
            *    14536    14635: gap of unknown length
            *    14636    16489: contig of 1854 bp in length
            *    16490    16589: gap of unknown length
            *    16590    19109: contig of 2520 bp in length
            *    19110    19209: gap of unknown length
            *    19210    21841: contig of 2632 bp in length
```

```
        *    21842    21941: gap of unknown length
        *    21942    23735: contig of 1794 bp in length
        *    23736    23835: gap of unknown length
        *    23836    25861: contig of 2026 bp in length
        *    25862    25961: gap of unknown length
        *    25962    29117: contig of 3156 bp in length
        *    29118    29217: gap of unknown length
        *    29218    31907: contig of 2690 bp in length
        *    31908    32007: gap of unknown length
        *    32008    36364: contig of 4357 bp in length
        *    36365    36464: gap of unknown length
        *    36465    43309: contig of 6845 bp in length
        *    43310    43409: gap of unknown length
        *    43410    48185: contig of 4776 bp in length
        *    48186    48285: gap of unknown length
        *    48286    55626: contig of 7341 bp in length
        *    55627    55726: gap of unknown length
        *    55727    65781: contig of 10055 bp in length
        *    65782    65881: gap of unknown length
        *    65882    69673: contig of 3792 bp in length
        *    69674    69773: gap of unknown length
        *    69774    79389: contig of 9616 bp in length
        *    79390    79489: gap of unknown length
        *    79490    91453: contig of 11964 bp in length
        *    91454    91553: gap of unknown length
        *    91554    96671: contig of 5118 bp in length
        *    96672    96771: gap of unknown length
        *    96772   100544: contig of 3773 bp in length
        *   100545   100644: gap of unknown length
        *   100645   111423: contig of 10779 bp in length
        *   111424   111523: gap of unknown length
        *   111524   115278: contig of 3755 bp in length
        *   115279   115378: gap of unknown length
        *   115379   124011: contig of 8633 bp in length
        *   124012   124111: gap of unknown length
        *   124112   138475: contig of 14364 bp in length
        *   138476   138575: gap of unknown length
        *   138576   140613: contig of 2038 bp in length
        *   140614   140713: gap of unknown length
        *   140714   144949: contig of 4236 bp in length
        *   144950   145049: gap of unknown length
        *   145050   164352: contig of 19303 bp in length
        *   164353   164452: gap of unknown length
        *   164453   170540: contig of 6088 bp in length
        *   170541   170640: gap of unknown length
        *   170641   185046: contig of 14406 bp in length
        *   185047   185146: gap of unknown length
        *   185147   222209: contig of 37063 bp in length
        *   222210   222309: gap of unknown length
        *   222310   253217: contig of 30908 bp in length.
FEATURES             Location/Qualifiers
     source          1..253217
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTD-3220F14"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    62246 a  62867 c  62532 g  61752 t   3820 others
ORIGIN

Alignment Scores:
Pred. No.:              0.175          Length:         253217
Score:                  186.50         Matches:        70
Percent Similarity:     47.98%         Conservative:   49
Best Local Similarity:  28.23%         Mismatches:     107
Query Match:            5.01%          Indels:         22
DB:                     2              Gaps:           6

US-10-008-355-2 (1-712) x AC016590 (1-253217)

Qy    483 AspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLys 502
              |||   :::          :::|||          ||||||   :::|||      |||
Db   4997 GATGGCATGTCGCAAGCGGACCTGGTTGCCGCCCTGGACAAACTGTATGCCGGCAGCAAG 4938

Qy    503 SerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAla---ValGluLeu 521
                ||||||   :::   |||      :::::::::   ||||||       :::
Db   4937 CTGGCCGACAAGGCCGAGCGCGCAGCCTGGCTGAAACGCACGCCTGCCGAATTCCAGGCC 4878

Qy    522 SerLys---SerValIleAla-AlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAl 540
          ||||||   |||   |||   ||||||  |||:::                     |||
Db   4877 AGCAAGGACAGCTTCATCCTAGGCCGCCGTCGCCATGTTACCCGGTATTCGCTGTAAC-- 4820

Qy    540 aTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrPro-- 559
                :::   |||   :::   |||               |||:::      |||
Db   4819 -AACGTAATGCGGAAGACTGAAGAACTGGCCGGTCAAGATCCACGCAAAGCTATGCCAAC 4761

Qy    560 -----------------------------------GlyArgAlaLeuProSerAspAl 567
                                             |||:::|||:::      |||||
Db   4760 TACATGAAAGCCAAGATCGCCTTCACTGAAGAGCAAGGGCCAAGCCGTCTATCCTGACGC 4701

Qy    567 aAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluPro---GlnAspGl 586
          ||||   |||:::|||:::::::::|||   :::   |||   :::          |||||
Db   4700 CAACAGCACCCTGCGCGTGACCTTCGGCCGCGTTGCCGGCCGTGACCATTGCGCCGACGG 4641

Qy    586 yAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys-S 606
          |      :::         ||||||      ||||||:::   |||          :::
Db   4640 CACCACCTGGACCGCGTTCACCACCGTCAATGGCGTGGTAGCCAAGGCCAGCGGGCGAAG 4581

Qy    606 erAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyA 626
             :::      |||      :::   :::                  |||::::::|||:
Db   4580 GCGAGTTCAATGCACCAGCGAAACAACTGGCCGCTGATCAATGGCCTAAGGACTTCGGCA 4521

Qy    626 rgTyrAlaGluAsn------GlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleT 644
          :::::   :::               :::   :::   :::|||:::      |||||||
Db   4520 AATTTGTCGATCCGAAGCTGAAAACCGTGCCGGTCGACTACCTGGCCACCCTGGACATCA 4461

Qy    644 hrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaP 664
          ||||||||||||||||||                 :::      |||      ||||||||||||
Db   4460 CCGGCGGCAACTCCGGTTCGGCTGCCCTGAACTCGAAAGGCGAATTCATCGGCCTGGCGT 4401

Qy    664 heAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluPro-AspLeuGlnArg 683
          ||||||||      :::::::::      |||   ::::::                |||
Db   4400 TCGACGGCACCCTGGACTCGATCATTTCCGACTGGGACTACAACAAGGCCGAACACCCGT 4341

Qy    684 ThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGlyGlnCysPro 703
          :::|||   ||||||:::||||||:::|||:::   :::
Db   4340 TCGATCCAGGTCGACCTGCGCTACATGCTGTGGAACATGAAACACGTCGACCACGCAGAC 4281

Qy    704 ArgLeuIleGlnGluLeu 709
             |||::::::|||:::
Db   4280 AACCTGCTGAAAGAAATG 4263

RESULT 3
AE011732
LOCUS       AE011732              10811 bp    DNA     linear   BCT 29-MAY-2002
DEFINITION  Xanthomonas axonopodis pv. citri str. 306,  section 110 of 469 of
            the complete genome.
ACCESSION   AE011732 AE008923
VERSION     AE011732.1  GI:21107161
KEYWORDS    .
SOURCE      Xanthomonas axonopodis pv. citri str. 306.
  ORGANISM  Xanthomonas axonopodis pv. citri str. 306
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xanthomonas.
REFERENCE   1  (bases 1 to 10811)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
```

```
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Comparison of the genomes of two Xanthomonas pathogens with
            differing host specificities
  JOURNAL   Nature 417 (6887), 459-463 (2002)
  MEDLINE   22022145
   PUBMED   12024217
REFERENCE   2  (bases 1 to 10811)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2001) Departmento de Bioquimica, Universidade de
            Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
            Brazil
FEATURES             Location/Qualifiers
     source          1. .10811
                     /organism="Xanthomonas axonopodis pv. citri str. 306"
                     /strain="306"
                     /db_xref="taxon:190486"
                     /note="pathovar: citri"
     gene            complement(293. .1315)
                     /gene="tdh"
                     /note="XAC1022"
     CDS             complement(293. .1315)
                     /gene="tdh"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="threonine 3-dehydrogenase"
                     /protein_id="AAM35905.1"
                     /db_xref="GI:21107162"
                     /translation="MKALVKREANKGIWLEQVPVPTPGPNEVLIKLEKTAICGTDLHI
                     YLWDEWSQRTIEPGLTIGHEFVGRVAELGSAVTGYQIGQRVSAEGHIVCGHCRNCRGG
                     RPHLCPNTVGIGVNVNGAFAEYMVMPASNLWPIPDQIPSELAAFFDPYGNAAHCALEF
                     DVIGEDVLITGAGPIGIIAAGICKHIGARNVVVTDVNDFRLKLAADMGATRVVNVSKT
                     SLKDVMADLHMEGFDVGLEMSGNSRAFNDMLDCMYHGGKIAMLGIMPRGAGCDWDKII
                     FKGLTVQGIYGRKMYETWYKMTQLVLSGFPLHKVLTHQLPIDDFQKGFDLMEEGKAGK
                     VVLSWN"
     gene            complement(1840. .4260)
                     /gene="fecA"
                     /note="XAC1023"
     CDS             complement(1840. .4260)
                     /gene="fecA"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="TonB-dependent receptor"
                     /protein_id="AAM35906.1"
                     /db_xref="GI:21107163"
                     /translation="MFHEPLTPCVAKRRLCTVRPPSAVPEQDSPMTSARRFHVTPLAL
                     ALATLLPVISQAATPADPAAVAPAEAALTADAAANATQLDAINVVSQGSTRQVQRISQ
                     QDIEQLTPGSSAFKAVEKLPGVQFQSADPFGTYEWSTQVTLHGFDQSRLGYTLDGIPL
                     GNMSYGVTNGLHITRAIISENLGSVEIAQGAGALGTASNTNLGGTMQFYSADPQTTPG
                     ARLVQTVGSDATRRTFVRADTGDRDGLSAYLSYANASTDKWKGYGDQTSEQANLKTVY
                     QWGDGNRLSLFVDTSRRKEYDYMDLSLTSQRALGWNYDYLQPDWATAVQMARAYQNTG
                     ATSGVANGYPQSLAGLPSDYSWLDASYYAGGGLRRDNLAGLSGTFVFGGATLDASGYY
                     HGNRGEGQWVTPYVRTSAQIPVSMRTTDYGLDRFGGTSALKWSWGNHDLEVGAWAENA
                     RTTQGRNYFALGNEGYTSLYNVYEAQTPFRRDFLQRYTTQTRMLYVQDTVRLLDDRLT
                     LNYGAKALSTTTRAQSLVPTTALAQGRIRAEDNFLPQLGINYKLDERQDLYASYSKNI
                     AAFGFTPFATSQAAFDRSRSTLEPEQSQTVQVGYRVQDAQFQLSADAYFTKFSNRLLT
                     TSPCTAVQTCASILSNVGAVHSAGADLALMWRPTEGLSWLNSLSWNRSRYQDDYLNNG
                     VVATSGKDVVGIPALMFSSSASYQIGNLRLDLDGKYVDKRYITFLNDSQVPSYWLFNA
                     GARYDFGRVGGVADVALALNISNLTDKRYFASTGTNGYVASDPDGYNQTMVVGAPRQY
                     FMSLDVKF"
     gene            4369. .6450
                     /gene="XAC1024"
     CDS             4369. .6450
                     /gene="XAC1024"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="non-hemolytic phospholipase C"
                     /protein_id="AAM35907.1"
                     /db_xref="GI:21107164"
                     /translation="MTSSSSRRDFLKRVAALTAAGALPSSIGRALALPANARTGTLRD
                     LEHVVILMQENRSFDHYFGALRGVRGFGDPRALQLRDGHAVWSQPTPDGRRVLPFAFD
                     SQNTCAPLIKSLDHSWKAGHGQDPARWAEYDAWVPYKGELTMGYFQRHDIPYYHALAD
                     AFTICDGYYCSLHGPTNPNRMYLFTGTSGPSVGNVGAQAVTNADDGNWTADMARDKPG
                     YAALQWTTYAQRLQAAGVDWRVYQEYDNFGCNSLAYFSHYRDLRTDDERYLRARACVP
                     GSTAGNAATTQAEHLVAAIAADVRANRLPQVSWVIPPTAYCEHPEAPPAYGESLVARL
                     IDALTANPEVWAKTALIINYDENDGFFDHVPAPLPALDARMGRSNVDTHGEVYDGVPI
                     GLGIRVPMLVISPWTRGGWVNSQVFDHTSVLRLLERRFGVAEPNISPWRRAVSGDLTS
                     VFDFRKPDDSALSALPSVDDYRARTAAVRDKPLPSAPAAATMPRQEPGQRPARALPYA
                     LQVHARVQEGAAVQLQFVNSGAAAAAFNVYNSAAGGGPWYYTVLPGTQLDDAPIGATH
                     EGAYALRVHGPNGFLREFAGQVPRASAPSAAPWVEARQEGDALVLEIGNAGQRACTVQ
                     LRALDYADPSARTLSLAAGQRETIRLALAASDHWYDLVVEQPGSAFRRRLAGHLETGR
                     PSRSDPAFGRV"
     gene            6617. .6766
                     /gene="XAC1025"
     CDS             6617. .6766
                     /gene="XAC1025"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAM35908.1"
                     /db_xref="GI:21107165"
                     /translation="MRPNLFAVSVLATLAVAASAQAGEGMWVPQQLPEIAGPPQQAGL
                     RLSPR"
     gene            6791. .7054
                     /gene="XAC1026"
     CDS             6791. .7054
                     /gene="XAC1026"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAM35909.1"
                     /db_xref="GI:21107166"
                     /translation="MGAVVALGGCTASFVSPQGLVVTNHHCAYGAIQLNSTAQKNLIK
                     DGFNAVRPADELSAGPSARIYVLDAITDVTAPAKAAMATPVRR"
     gene            7051. .7236
                     /gene="XAC1027"
     CDS             7051. .7236
                     /gene="XAC1027"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAM35910.1"
                     /db_xref="GI:21107167"
                     /translation="MMDAQGKLVGLAFDGNWESVSSNWIFDPAMTRMIAVDGRYLRWI
                     MTEVAPAPQLLKELGVR"
```

```
     gene            7329..7973
                     /gene="pgmA"
                     /note="XAC1028"
     CDS             7329..7973
                     /gene="pgmA"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoglycerate mutase"
                     /protein_id="AAM35911.1"
                     /db_xref="GI:21107168"
                     /translation="MRILLARHGETPWNAEGRYQGQIDIPLSPVGEGQARALGERLHS
                     LQIDRAVASPLSRAQATAKAALGASRQALLQTDADLQEIAHGEWEGLLASEINDKDPA
                     RLRAWREEPDTVLMPGGESLRQVLDRSWRGLMRAADGLGAHDTLLVVAHDAVNRVILC
                     KILGLPLSRLWSFRQAPTTLNLLEGDDVEHLEVVRLNDCAHHTPFFGEAKHRAL"
     gene            8242..9540
                     /gene="folC"
                     /note="XAC1029"
     CDS             8242..9540
                     /gene="folC"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark; dihydrofolate
                     synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="folylpolyglutamate synthase"
                     /protein_id="AAM35912.1"
                     /db_xref="GI:21107169"
                     /translation="MNATDSLSAWLAYIEQQHPSAIAMGLERVREVAVRLQIAAPAKH
                     VIVVGGTNGKGSTVAFIEAIGRAAGWKVGAYTSPHLLRYNERVRIDGQEASDAQLVDA
                     FAAVEAARGDTALTYFEFGTLAALWLFQQSALELAVLEIGLGGRLDAVNIVDSDVAVI

Alignment Scores:
Pred. No.:              0.0237        Length:        10811
Score:                  175.00        Matches:       32
Percent Similarity:     75.00%        Conservative:  16
Best Local Similarity:  50.00%        Mismatches:    16
Query Match:            4.71%         Indels:        0
DB:                     1             Gaps:          0

US-10-008-355-2 (1-712) x AE011732 (1-10811)

Qy    646 GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAsp 665
          ||||||||||||||||||    |||       |||:::|||:::||||||||||||||
Db   7033 GGCAACTCCGGTTCGCCGGTGATGGACGCGCAGGGCAAGCTGGTCGGCCTGGCGTTCGAC 7092

Qy    666 GlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 685
          |||||||||||:::::::|||    :::      |||:::|||    :::    |||   |||
Db   7093 GGAAATTGGGAGTCGGTGAGCAGCAACTGGATCTTCGACCCGGCAATGACCCGCATGATC 7152

Qy    686 SerValAspIleArgTyrValLeuPheMetIleAspLysTrpGlyGlnCysProArgLeu 705
          :::||||||  ||||||:::   :::::::::    :::             |||:::|||
Db   7153 GCCGTGGACGGCCGCTATCTGCGTTGGATCATGACCGAGGTCGCTCCGGCGCCACAGCTG 7212

Qy    706 IleGlnGluLeu 709
          ::::::||||||
Db   7213 TTGAAGGAACTC 7224

RESULT 4
AE014150
LOCUS       AE014150               50372 bp    DNA     linear   BCT 18-JUL-2002
DEFINITION  Streptococcus pyogenes MGAS315, section 15 of 37 of the complete
            genome.
ACCESSION   AE014150 AE014074
VERSION     AE014150.1  GI:21904382
KEYWORDS    .
SOURCE      Streptococcus pyogenes MGAS315.
  ORGANISM  Streptococcus pyogenes MGAS315
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 50372)
  AUTHORS   Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
            Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
            Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
            Musser,J.M.
  TITLE     Genome sequence of a serotype M3 strain of group A Streptococcus:
            Phage-encoded toxins, the high-virulence phenotype, and clone
            emergence
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
REFERENCE   2  (bases 1 to 50372)
  AUTHORS   Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
            Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
            Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
            Musser,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
            Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
            Hamilton, MT 59840, USA
FEATURES             Location/Qualifiers
     source          1..34426
                     /organism="Streptococcus pyogenes MGAS315"
                     /strain="MGAS315"
                     /serotype="M3"
                     /db_xref="taxon:198466"
                     /focus
                     /note="group: A"
     source          34427..>50372
                     /organism="Streptococcus pyogenes phage 315.1"
                     /db_xref="taxon:198538"
     gene            208..1206
                     /gene="SpyM3_0653"
     CDS             208..1206
                     /gene="SpyM3_0653"
                     /note="best non-GAS blastp hit: gb|AAK99779.1| (AE008470)
                     ABC transporter substrate-binding protein - unknown
                     substrate [Streptococcus pneumoniae R6]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ABC transporter substrate-binding
                     protein"
                     /protein_id="AAM79260.1"
                     /db_xref="GI:21904383"
                     /translation="MKNKSLIATLLVLTVIVIGSLLSKGVSKENRDLANQPNITIGIL
                     QFVTHEALDDIKRGIEDQLKKQMPQKQNVVIKVMNAEGDQSKIQTMSRQLVQSGSDIV
                     IGIATPAAQGLAAASKDIPVVMSAVSDPVGSRLVMQLDQPEANVTGLSNKVPVKQTID
                     LMKKLTPHVKTVGILYASNEDNSLSQVKEFRRLAKKKGYQVISYAVPSTNEVPATMSV
                     MLGKVDAVFIPQDNTIASAFSSVMTTSKAAKIPVYTSVDRMVEKGGLAAISQNQYDLG
                     VQTANQVLKLIKGKRVVDVPVKVVDIGQPLINKNVAAELGIAINREDFPSASFIEN"
     gene            1157..1375
                     /gene="SpyM3_0654"
     CDS             1157..1375
                     /gene="SpyM3_0654"
                     /note="best blastp hit: gb|AAK33913.1| (AE006546)
                     hypothetical protein [Streptococcus pyogenes M1 GAS]"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAM79261.1"
                     /db_xref="GI:21904384"
                     /translation="MLLTGKIFLQLLLRINTKTCDIKGQEKIIKNTLTEMNLDAKSD
                     NRLEDVTVIFIIFSFKMLYFRHVKNLAF"
     gene            1551..2420
                     /gene="SpyM3_0655"
     CDS             1551..2420
                     /gene="SpyM3_0655"
                     /note="best non-GAS blastp hit: gb|AAK99780.1| (AE008470)
                     ABC transporter membrane-spanning permease - unknown
                     substrate [Streptococcus pneumoniae R6]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ABC transporter membrane-spanning
                     permease"
                     /protein_id="AAM79262.1"
                     /db_xref="GI:21904385"
                     /translation="MIVSSVSQGLIWGVLGLGIYLTFRILNFPDMTTEGSFPLGGAVA
```

```
                     VTAISLGWNPFLSTLLGMLSGALAGFLTGLLYTKGKMPTLLAGILVMTSCNSIMLMVM
                     GRANLGLHDHKRIQDCLPFSIDLNSLLTGLITVVIVISVLIYFLYTNLGQAYIATGDN
                     KDMAKSFGINTDWMEVMGLVVSNSLIALSGALVSQQDGYADVSKGIGVIVIGLASIIV
                     GEVLYSTGLTLLERLIAIVIGSILYQFLISVVIMLGFNTSYLKLISALVLALCLMIPV
                     VKERFFKGVRLTR"
     gene            2417. .3175
                     /gene="SpyM3_0656"
     CDS             2417. .3175
                     /gene="SpyM3_0656"
                     /note="best blastp hit: gb|AAK33915.1| (AE006547) putative
                     ABC transporter (ATP-binding protein) [Streptococcus
                     pyogenes M1 GAS]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ABC transporter (ATP-binding protein)"
                     /protein_id="AAM79263.1"
                     /db_xref="GI:21904386"
                     /translation="MTKIIELINATVDVDNGFEDAKTILDNVTLTIYEHDFLTILGGN
                     GAGKSTLFNVIAGTLSLTRGQIRILGQDVTYWPAEKRALYLSRVFQDPKMGTAPRMTV
                     AENLLIARQRGEKRSLASRKITEHLASFEDLVKRTGNGLEKHLETPAGLLSGGQRQAL
                     SLLMATLKKPALLLLDEHTAALDPKTSHSLMQLTDEFVTKDGLTALMITHHMEDALTY
                     GNRLIVMKDGNIIKDLNQMEKEQLTITDYYQLFD"
     gene            3379. .5040
                     /gene="SpyM3_0657"
     CDS             3379. .5040
                     /gene="SpyM3_0657"
                     /note="best non-GAS blastp hit: gb|AAK74765.1| (AE007370)
                     metallo-beta-lactamase superfamily protein [Streptococcus
                     pneumoniae TIGR4], andgb|AAK99342.1| (AE008432) Conserved
                     hypothetical protein [Streptococcus pneumoniae R6]"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAM79264.1"
                     /db_xref="GI:21904387"
                     /translation="MTDIKMIALGGVREYGKNFYLVEINDSMFILDAGLKYPENEQLG
                     VDLVIPNLDYVIENKGKVQGIFLSHGHADAIGALPYLLAEVSAPVFGSELTIELAKLF
                     VKSNNSTKKFNNFHVVDSDTEIEFKDGLVSFFRTTHSIPESMGIVIGTDKGNIVYTGD
                     FKFDQAAREGYQTDLLRLAEIGKEGVLALLSDSVNATSNDQIASESEVGEEMDSVISD
                     ADGRVIVAAVASNLVRIQQVFDSATAHGRRVVLTGTDAENIVRTALRLEKLMITDERL
                     LIKPKDMSKFEDHELIILEAGRMGEPINSLQKMAAGRHRYVQIKEGDLVYIVTTPSTA
                     KEAMVARVENLIYKAGGSVKLITQNLRVSGHANGRDLQLLMNLLKPQYLFPVQGEYRD
                     LAAHAKLAEEVGIFPENIHILKRGDIMVLNDEGFLHEGGVPASDVMIDGNAIGDVGNI
                     VLRDRKVLSEDGIFIVAITVSKKEKRIISKAKVNTRGFVYVKKSHDILRESAELVNTT
                     VGNYLKKDTFDWGELKGNVRDDLSKFLFEQTKRRPAILPVVMEVR"
     gene            5173. .5958
                     /gene="estA"
                     /note="SpyM3_0658"
     CDS             5173. .5958
                     /gene="estA"
                     /note="best blastp hit: gb|AAK33917.1| (AE006547) putative
                     tributyrin esterase [Streptococcus pyogenes M1 GAS]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tributyrin esterase"
                     /protein_id="AAM79265.1"
                     /db_xref="GI:21904388"
                     /translation="MASIAIEYHSVVLGMERKVNVIYPDQSEIPKKDQGDKDIPVLYL
                     LHGMGGNENSWQKRTAIERLLRHTNLIVVMPSTDLGWYTDTAYGLNYYRALSQELPQV
                     LAAFFPNMTQKREKTFVAGLSMGGYGAFKWALKSNRFSYAASFSGALDFSPETLLEGN
                     LGELAYWQGVFGQFDDPDLDKHYLKNMVAESDGKTKFYAWCGYEDFLFATNEKAIADF
                     QAQGLDIDYHKGHGKHEWYYWNQQLEVLLEWLPINYQKEERLS"
     gene            5988. .6314
                     /gene="SpyM3_0659"
     CDS             5988. .6314
                     /gene="SpyM3_0659"
                     /note="best blastp hit: gb|AAK33918.1| (AE006547)
                     hypothetical protein [Streptococcus pyogenes M1 GAS]"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAM79266.1"
                     /db_xref="GI:21904389"
                     /translation="MKQLLAIILWLPKLIVKMFWHLIKGFLQTILLVTIIIIGLMYYS
                     NHSDSVLANKISIVTEQVVQIFDILTQKPSAKTRHGSGNSHSTFLVSAKEGNQRLKKS
                     SISFGN"
     gene            6360. .8267
                     /gene="SpyM3_0660"
     CDS             6360. .8267
                     /gene="SpyM3_0660"
                     /note="best blastp hit: gb|AAK33919.1| (AE006547) putative
                     ABC transporter (ATP-binding protein) [Streptococcus
                     pyogenes M1 GAS]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ABC transporter (ATP-binding protein)"
                     /protein_id="AAM79267.1"
                     /db_xref="GI:21904390"
                     /translation="MIILQGNKLERSFSGDVLFQNISLQVDERDRIALVGPNGAGKST
                     LLKLLVGEETPTSGEVNTKKDLTLSYLAQNSRFESDKTIYEEMLKVFEALRQDEKRLR
                     QMEMDMATVSGQDLTRLMMDYDLLAERFRQQGGFTYEAEIKAILNGFKFDESMWQMTI
                     AELSGGQNTRLALAKMLLEKPELLVLDEPTNHLDIETIAWLENYLANYQGALIIVSHD
                     RYFLDKVATVTLDLTPHGLDRYVGNYSRFMTLKAEKLVAEEKQFDKQQKEIAKLEDFV
                     QKNIVRASTTKRAQARRKQLEKIERLDKPTSARKSAHMTFQADKPSGNVVLTVEKAAI
                     GYNQHVLSEPINLDVHKLDAIAIVGPNGIGKSTLIKSVIGQIPFIKGEVKYGANVEIG
                     YYDQTQSHLTSSNTVLEELWQDFSTTPEVDIRNRLGAFLFSGDDVKKSVAMLSGGEKA
                     RLLLAKLSMENNNFLVLDEPTNHLDIDSKEVLENALIDFDGTLLFVSHDRYFINRLAT
                     KVLEITENGSTLYLGDYDYYLEKKAELEELARLAAGETVEETKEASATDYQLQKANQK
                     ERRRLTRRYEEIEARLETIEERIGAIQEDMHASNDTAQLIAWQKEWDQLDQEQEALME
                     EWETIAEQIES"
     gene            8552. .9520
                     /gene="acoA"
                     /note="SpyM3_0661"
     CDS             8552. .9520
                     /gene="acoA"
                     /note="best blastp hit: gb|AAK33920.1| (AE006547) putative
                     acetoin dehydrogenase (TPP-dependent) alpha chain
                     [Streptococcus pyogenes M1 GAS]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative acetoin dehydrogenase (TPP-dependent)
                     alpha chain"
                     /protein_id="AAM79268.1"
                     /db_xref="GI:21904391"
                     /translation="MVTVSKEQHLDMFLKMERIREFDSRINKLVRRGFVQGMTHFSVG
```

```
Alignment Scores:
Pred. No.:              12.4        Length:         50372
Score:                  147.50      Matches:        171
Percent Similarity:     32.47%      Conservative:   105
Best Local Similarity:  20.12%      Mismatches:     232
Query Match:            3.97%       Indels:         343
DB:                     1           Gaps:           44

US-10-008-355-2 (1-712) x AE014150 (1-50372)

Qy      35 AsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeu 54
           |||:::||||||   |||   :::   ::::::      ||||||   |||
Db   30183 AATGAAGAAAACGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAA--- 30239

Qy      55 TyrSerPheAspLysProSerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThr 74
                    |||:::      |||         :::      :::
Db   30240 ---------GATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCTCAC--------- 30281

Qy      75 GlyIleThrValSerAspGlnGlyLeu------IlePheThrAsnHisHisCysGlyTyr 92
                          |||:::   :::      :::            :::   |||:::
Db   30282 ------CTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGG 30335

Qy      93 GlyAlaIle---------------------------------GlnSerGlnSerThrVal 101
           |||   :::                                 ||||||      |||:::
Db   30336 GGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATA 30395

Qy     102 AspHisAspTyrLeuArg---AspGlyPheValSerArg--------------------- 113
                 |||:::|||:::   |||||||||   :::|||
Db   30396 ---TTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCAT 30452
```

```
Qy    114 ------------ThrMetGlyGluGluLeuProIleProGlyLeuSer------------ 125
                      |||    |||:::::               :::|||
Db  30453 GATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGAT 30512

Qy    126 ---------------------------------------------ValLysTyrLeuArg 130
                                                       |||    :::
Db  30513 AGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTA 30572

Qy    131 LysIleValLysValThrAsp---------LysValGluGlyGln--------------- 142
          :::  |||||||||   |||         ||||||   |||:::
Db  30573 CAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGCGGCATAAGCCAGAAAAT 30632

Qy    143 ------------------------------------LeuLysGlyIleThrAspGluMet 150
                                              ||||||       :::
Db  30633 ATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGT 30692

Qy    151 GluArgLeuArgLysAlaGlnGluValCysGlnGluLeuAlaLys---------LysGlu 167
          ||||||:::::::::   :::|||         :::||||||         ||||||
Db  30693 GAGCGTATGAAACGTATTGAAGAAGGAATAAAAGAACTAGGAAGTGATATTCTAAAGGAG 30752

Qy    168 AsnAlaAspGluAsn---GlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyr 186
                   ||||||   ||||||                            ||||||
Db  30753 TATCCTGTTGAAAACACTCAATTA------------------------CAAAATGAAAAG 30788

Qy    187 PheLeuIleValTyr---------------AspValPheLysAspValArgMetValPhe 201
             :::   |||               |||::::::   |||         :::
Db  30789 CTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATT 30848

Qy    202 AlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArg 221

Db  30848 ------------------------------------------------------------ 30848

Qy    222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGlu 241
                                                       ||||||   ::::::
Db  30849 ---------------------------------------------AATCGTTTAAGTGAT 30863

Qy    242 TyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGly 261
          |||      |||:::                              :::      |||
Db  30864 TATGATGTCGATCAC------------------------------ATTGTTCCACAAAGT 30893

Qy    262 TyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu 281
          :::      ||||||                           |||   |||      :::
Db  30894 TTCCTTAAAGACGAT---------------------------TCAATAGACAATAAGGTC 30926

Qy    282 ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArg 301
             :::         :::::|||   :::::::::|||   |||      ||||||
Db  30927 TTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTC 30986

Qy    302 GlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLys 321
                   :::      |||:::::      ::::::|||         |||:::   |||
Db  30987 AAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAG 31046

Qy    322 TyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArg 341
          :::                           |||      |||         |||
Db  31047 TTT---------------------GATAATTTAACAAAAGCTGAACGTGGA--------- 31076

Qy    342 GlyLeuAlaArgLeuAspValIleGly---------ArgLysArgAlaGluGluArgAla 358
          ||||||:::   ||||||   :::|||         |||:::      |||   |||
Db  31077 GGTTTGAGTGAACTTGATAAAGTTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAA 31136

Qy    359 PheAlaAspTrpIle------------------ArgLysAsnGlyLys------------ 368
                   :::                     |||   |||   |||
Db  31137 ATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAAT 31196

Qy    368 ------------------------------------------------------------ 368

Db  31197 GATAAACTTATTCGAGAGGTTAGAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTC 31256

Qy    368 ------------------------------------------------------------ 368

Db  31257 CGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGAT 31316

Qy    368 ------------------------------------------------------------ 368

Db  31317 GCGTATCTTAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCG 31376

Qy    369 SerAlaValTyrGly---------------AspValLeuSerSerLeuGluLysAlaTyr 383
             |||||||||               ||||||         :::   |||:::
Db  31377 GAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAG 31436

Qy    384 LysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu---PheGly 402
          :::|||      ||||||      :::      :::   |||      :::   |||
Db  31437 CAGGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTC 31496

Qy    403 GlyThrGluVal----ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAs 421
             ||||||:::   ::::::      :::||||||||||||||||             ::
Db  31497 AAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTA-ATCGAAACTAA 31555

Qy    421 pAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLe 441
          :          |||      :::   :::   ||||||   :::|||:::      :::
Db  31556 TGGGGAAACTGGA------GAAATTGTCTGGGATAAAGGGCGAGATTTT------GCCAC 31603

Qy    441 uAspArgLysValLeuPro---AlaMetLeuAspIleValArgArgArg---------- 456
             ||||||||||||         :::::::||||||::::::
Db  31604 AGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGAC 31663

Qy    457 ------------------------IlePro---------AlaAspLysLeuProAspIl 465
                                  :::|||         :::|||||||||
Db  31664 AGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTT-------- 31715

Qy    465 ePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPheVa 485
                      |||   :::|||         |||   |||||||||
Db  31716 ------------ATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGT----- 31757

Qy    485 lPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLysSerMetAs 505
           ||||||         |||   ||||||         |||   ::::::
Db  31758 -TTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAA---GG 31813

Qy    505 pLysGluLysPheAlaLysAlaIle----------------------------------- 513
           |||   |||      |||::::::
Db  31814 GAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATAATGGAAAGAAG 31873

Qy    514 -------GluLysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAl 531
                 ||||||||||||      |||                        :::
Db  31874 CTCTTTTGAAAAAGATCCGATTGACTTTTAGAAGCTAAA---------GGATATAAGGA 31924

Qy    531 aIleGlnAlaAspAlaMetAla--------------AsnAlaTyrAlaIleGluLysGl 546
           ::::::   |||   :::                 :::   :::   :::|||   ||
Db  31925 AGTTAGAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGG 31984

Qy    546 yLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArg------AlaLeuPr 564
          |::::::            |||         |||:::      |||        |||||||
Db  31985 TCGTAAACGGATGCTGGCTAGTGCCGGAGAATTGCAAAAAGGAAATGAGCTAGCTCTGCC 32044

Qy    564 oSerAsp---AlaAsnPheThrMetArgMetSer---TyrGlySerIleLysGlyTyrGl 582
          ||||         ||||||            |||   |||         :::||||||
Db  32045 AAGCAAATATGTGAATTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGT---AG 32101

Qy    582 uProGlnAspGly-----------AlaTrpTyrAsnTyrHisThrThrGlyLysGlyVa 598
           |||:::|||                        :::      |||      |||
Db  32102 TCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCAT---------AAGCATTA 32152

Qy    598 lLeuGluLysGlnAspProLysSerAspGluPheAlaValGln----------------- 612
           |||::::::         :::      ||||||:::   :::
Db  32153 TTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGC 32212

Qy    612 ------------------------------------------------------------ 612

Db  32213 CAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACA 32272

Qy    613 ----GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly----------------- 625
                |||||||||:::   ||||||         |||   |||
Db  32273 AGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAA 32332

Qy    626 --------------------------ArgTyrAlaGluAsnGlyGlnLeuHisIleAl 636
                                    ||||||            ::::::      ||
```

```
Db  32333 ATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGC 32392

Qy    636 aPheLeuSerAsnAsnAspIleThrGly 645
          |   |||·  :::      |||||||||
Db  32393 CACTCTTATCCATCAATCCATCACTGGT 32420

RESULT 5
AE006549
LOCUS       AE006549              10889 bp    DNA     linear   BCT 01-JUN-2001
DEFINITION  Streptococcus pyogenes M1 GAS strain SF370, section 78 of 167 of
            the complete genome.
ACCESSION   AE006549 AE004092
VERSION     AE006549.1  GI:13622187
KEYWORDS    .
SOURCE      Streptococcus pyogenes M1 GAS.
  ORGANISM  Streptococcus pyogenes M1 GAS
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 10889)
  AUTHORS   Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
            Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
            Qian,Y., Jia,H.G., Najar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
            Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
  TITLE     Complete genome sequence of an M1 strain of Streptococcus pyogenes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
  MEDLINE   21192684
   PUBMED   11296296
REFERENCE   2  (bases 1 to 10889)
  AUTHORS   Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
            Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
            Qian,Y., Jia,H.G., Najar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
            Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2001) Department of Microbiology and Immunology,
            University of Oklahoma Health Sciences Center, 940 SL Young Blvd,
            Oklahoma City, OK 73104, USA
FEATURES             Location/Qualifiers
     source          1..10889
                     /organism="Streptococcus pyogenes M1 GAS"
                     /strain="SF370"
                     /serotype="M1"
                     /db_xref="taxon:160490"
     gene            98..739
                     /gene="SPy1039"
     CDS             98..739
                     /gene="SPy1039"
                     /note="Best Blastp hit = no blast hits"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK33931.1"
                     /db_xref="GI:13622188"
                     /translation="MTKQQIEKLADLLPPFAATLYLDMIPEFQEGYASYVYQTPKQDI
                     QKRRIKQLTKNFERLLDTPSFYVMPVTRKTAEQLLADKGLAYHLLSTVAANQVFAAAV
                     NEDQMLFQVLRNGARVAFVILQEKGEFLEMTLGILETVTENEVPAAIVEAVQTYLLAD
                     MQEKLPSLSIMVSQESSKLFYQEQGFVLLAKGCKPGNENDSDVCQLVKYFSKP"
     gene            802..1932
                     /gene="hemN"
                     /note="SPy1040"
     CDS             802..1932
                     /gene="hemN"
                     /EC_number="1.3.3.3"
                     /note="Best Blastp hit = pir||B69640 coproporphyrinogen
                     oxidase (EC 1.3.3.3) III, oxygen-independent hemN -
                     Bacillus subtilis >gi|1303805|dbj|BAA12461.1| (D84432)
                     YqeR [Bacillus subtilis] >gi|2634996|emb|CAB14492.1|
                     (Z99117) coproporphyrinogen III oxidase [Bacillus
                     subtilis]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative coproporphyrinogen III oxidase"
                     /protein_id="AAK33932.1"
                     /db_xref="GI:13622189"
                     /translation="MSKKPTSAYVHIPFCTQICYYCDFSKVFIQNQPVDAYLKALIQE
                     FDSYGIRDLKTLYIGGGTPTAITAKQLEYLLNHLERNLNLDDLEEFTIEANPGDLTPE
                     KIAVLQRSAVNRISLGVQTFNNKQLKQIGRSHNEEQIYSTIANLKTAGFHNISIDLIY
                     ALPGQTLDQVKENVAKALALDIPHLSLYSLILEHHTVFMNKMRRGKLNLPTEDLEAEM
                     FEYIISEMEANGFEHYEISNFTKPGFESRHNLMYWDNVEYFGCGAGASGYLNGIRYQN
                     RVPIQHYLKAVEAGNARLNEEVLRKEEMMEEELFLGLRKKTGVSIQRFQEKFGMSFEE
                     RYGNIVRELQNQGLLVKDDAFVRMTKKGLFLGDSVAERFILD"
     gene            2126..2695
                     /gene="SPy1042"
     CDS             2126..2695
                     /gene="SPy1042"
                     /note="Best Blastp hit = gb|AAB71730.1| (U65643) acyl-ACP
                     thioesterase [Myristica fragrans]"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK33933.1"
                     /db_xref="GI:13622190"
                     /translation="MRSLFIACHIFKKPSPLKQKPFPIINFFVIANFIFMIKRGGLLV
                     DILAYFALLNPDTRKVATIPEDLVAPFETDFVKKLHRVPKMPLLEQSIDRDYYVRYFD
                     IDMNGHVNNSKYLDWMYDVLGCEFLKTHQPLKMTLKYVKEVSPGGQITSSYHLDQLTS
                     YHQITSDGQLNAQAMIEWRAIKQTESETD"
     gene            2695..3459
                     /gene="SPy1043"
     CDS             2695..3459
                     /gene="SPy1043"
                     /note="Best Blastp hit = pir||H70023 N-acetyl-glucosamine
                     catabolism homolog yutF - Bacillus subtilis
                     >gi|2635726|emb|CAB15219.1| (Z99120) similar to
                     N-acetyl-glucosamine catabolism [Bacillus subtilis]"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK33934.1"
                     /db_xref="GI:13622191"
                     /translation="MPYKGYLIDLDGTIYQGKNRIPAGERFIKRLQERGIPYLLVTNN
                     TTRTPEMVQSMLANQFHVETSIETIYTATMATVDYMNDMNRGKTAYVIGETGLKSAIA
                     AAGYVEELENPAYVVVGLDSQVTYEMLAIATLAIQKGALFIGTNPDLNIPTERGLMPG
                     AGALNALLEAATRVKPVFIGKPNAIIMNKSLEVLGIQRSEAVMVGDNYLTDIMAGIQN
                     DIATILVTTGFTRPEEVPTLPIQPDHVLSSLDEWRL"
     gene            3459..4091
                     /gene="SPy1044"
     CDS             3459..4091
                     /gene="SPy1044"
                     /note="Best Blastp hit = sp|P54589|YHCE_BACSU HYPOTHETICAL
                     29.5 KD PROTEIN IN GLPD-CSPB INTERGENIC REGION
                     >gi|7475050|pir||A69822 hypothetical protein yhcE -
                     Bacillus subtilis >gi|1239980|emb|CAA65688.1| (X96983)
                     hypothetical protein [Bacillus subtilis]
                     >gi|2633228|emb|CAB12733.1| (Z99108) yhcE [Bacillus
                     subtilis]"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK33935.1"
                     /db_xref="GI:13622192"
                     /translation="MVENTKLLCSWVWLLALAILITIYSTWLWYPLEVDHLKLEQVVF
                     MSKDAILHNYNGLLNYLTNPFVTRLEFANFHSSADGLKHFADVKWLFHLTQVVFLGLL
                     YPTLKTFTQRLKTKRFWLLQKPLILAALFPLMIGLMASFIGFEHFFTLFHQVLFVGDS
                     SWLFDPLKDSVIWILPEVFFLHCFLFFMIVYEIILWSLVGLARWQRLNVE"
     gene            4571..8677
                     /gene="SPy1046"
     CDS             4571..8677
                     /gene="SPy1046"
                     /note="Best Blastp hit = pir||H81982 hypothetical protein
                     NMA0631 [imported] - Neisseria meningitidis (group A
                     strain Z2491) >gi|7379363|emb|CAB83920.1| (AL162753)
                     hypothetical protein NMA0631 [Neisseria meningitidis]"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK33936.1"
                     /db_xref="GI:13622193"
```

```
                    /translation="MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK
                    KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
                    ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH
                    MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSK
                    SRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD
                    NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL
                    LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK
                    LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
                    YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
                    PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
                    DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
                    LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
                    LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
                    VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKE
                    HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
                    LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
                    GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
                    YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
                    KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
                    MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
                    AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF
                    ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
                    KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
                    PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD"
     gene           8677..9546
                    /gene="SPy1047"
     CDS            8677..9546
                    /gene="SPy1047"
                    /note="Best Blastp hit = pir||E81299 hypothetical protein
                    Cj1522c [imported] - Campylobacter jejuni (strain NCTC
                    11168) >gi|6968946|emb|CAB73942.1| (AL139078) hypothetical
                    protein Cj1522c [Campylobacter jejuni]"
                    /codon_start=1
                    /transl_table=11
                    /product="hypothetical protein"
                    /protein_id="AAK33937.1"
                    /db_xref="GI:13622194"
                    /translation="MAGWRTVVVNTHSKLSYKNNHLIFKDAYKTELIHLSEIDILLLE
                    TTDIVLSTMLVKRLVDENVLVIFCDDKRLPTAMLMPFYGRHDSSLQLGKQMSWSETVK
                    SQVWTTIIAQKILNQSCYLGACSYFEKSQSIMDLYHGLENFDPSNREGHAARIYFNTL
                    FGNDFSRDLEHPINAGLDYGYTLLLSMFAREVVVSGCMTQFGLKHANQFNQFNFASDI
                    MEPFRPLVDKIVYENRNQPFPKIKRELFTLFSDTFSYNGKEMYLTNIISDYTKKVVKA
                    LNNEGKGVPEFRI"
     gene           9543..9884
                    /gene="SPy1048"
     CDS            9543..9884
                    /gene="SPy1048"
                    /note="Best Blastp hit = emb|CAB83918.1| (AL162753)
                    hypothetical protein NMA0629 [Neisseria meningitidis]"
                    /codon_start=1
                    /transl_table=11
                    /product="hypothetical protein"
                    /protein_id="AAK33938.1"
                    /db_xref="GI:13622195"
                    /translation="MSYRYMRMILMFDMPTDTAEERKAYRKFRKFLLSEGFIMHQFSI
                    YSKLLLNNTANNAMIGRLREHNPNKGNITLLTVTEKQFARMIYLHGERNNCIANSDER
                    LVFLGEAFDES"
     gene           9874..10536
                    /gene="SPy1049"
     CDS            9874..10536
                    /gene="SPy1049"
                    /note="Best Blastp hit = pir||T03910

Alignment Scores:
Pred. No.:              2.26        Length:          10889
Score:                  146.50      Matches:         172
Percent Similarity:     31.98%      Conservative:    103
Best Local Similarity:  20.00%      Mismatches:      223
Query Match:            3.94%       Indels:          363
DB:                     1           Gaps:            44

US-10-008-355-2 (1-712) x AE006549 (1-10889)
```

```
Qy       35 AsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeu 54
Db     6395 AATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAA--- 6451

Qy       55 TyrSerPheAspLysProSerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThr 74
Db     6452 ---------GATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCTCAC--------- 6493

Qy       75 GlyIleThrValSerAspGlnGlyLeu------IlePheThrAsnHisHisCysGlyTyr 92
Db     6494 ------CTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGG 6547

Qy       93 GlyAlaIle---------------------------------GlnSerGlnSerThrVal 101
Db     6548 GGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATA 6607

Qy      102 AspHisAspTyrLeuArg---AspGlyPheValSerArg--------------------- 113
Db     6608 ---TTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCAT 6664

Qy      114 ------------ThrMetGlyGluGluLeuProIleProGlyLeuSer------------ 125
Db     6665 GATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGAT 6724

Qy      126 -------------------------------------------ValLysTyrLeuArg 130
Db     6725 AGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTA 6784

Qy      131 LysIleValLysValThrAsp---------LysValGluGlyGln--------------- 142
Db     6785 CAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAAT 6844

Qy      143 -----------------------------------LeuLysGlyIleThrAspGluMet 150
Db     6845 ATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGA 6904

Qy      151 GluArgLeuArgLysAlaGlnGluValCysGlnGluLeuAlaLys---------LysGlu 167
Db     6905 GAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAG 6964

Qy      168 AsnAlaAspGluAsn---GlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyr 186
Db     6965 CATCCTGTTGAAAATACTCAATTG-----------------------CAAAATGAAAAG 7000

Qy      187 PheLeuIleValTyr--------------AspValPheLysAspValArgMetValPhe 201
Db     7001 CTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATT 7060

Qy      202 AlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArg 221
Db     7060 ------------------------------------------------------------ 7060

Qy      222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGlu 241
Db     7061 ---------------------------------------------AATCGTTTAAGTGAT 7075

Qy      242 TyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGly 261
Db     7076 TATGATGTCGATCAC-----------------------------ATTGTTCCACAAAGT 7105

Qy      262 TyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu 281
Db     7106 TTCCTTAAAGACGAT--------------------------TCAATAGACAATAAGGTC 7138

Qy      282 ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArg 301
Db     7139 TTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTC 7198

Qy      302 GlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLys 321
Db     7199 AAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAG 7258

Qy      322 TyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArg 341
```

```
                 :::                    |||        |||        |||
Db     7259 TTT--------------------GATAATTTAACGAAAGCTGAACGTGGA--------- 7288

Qy      342 GlyLeuAlaArgLeuAspValIleGly---------ArgLysArgAlaGluGluArgAla 358
             ||||||:::  |||||         |||        |||:::      |||   |||
Db     7289 GGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAA 7348

Qy      359 PheAlaAspTrpIle-----------------ArgLysAsnGlyLys------------ 368
                        :::                |||  |||  |||
Db     7349 ATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAAT 7408

Qy      368 ------------------------------------------------------------ 368

Db     7409 GATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTC 7468

Qy      368 ------------------------------------------------------------ 368

Db     7469 CGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGAT 7528

Qy      368 ------------------------------------------------------------ 368

Db     7529 GCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCG 7588

Qy      369 SerAlaValTyrGly--------------AspValLeuSerSerLeuGluLysAlaTyr 383
                |||||||||              ||||||        :::   |||:::
Db     7589 GAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAG 7648

Qy      384 LysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu---PheGly 402
             :::|||     ||||||     :::     :::   |||     :::   |||
Db     7649 CAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTC 7708

Qy      403 GlyThrGluVal----ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAs 421
               ||||||:::    ::::::     :::||||||||||||||           ::
Db     7709 AAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTA-ATCGAAACTAA 7767

Qy      421 pAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLe 441
             :          |||     :::   :::  ||||||   :::|||:::       :::
Db     7768 TGGGGAAACTGGA------GAAATTGTCTGGGATAAAGGGCGAGATTTT------GCCAC 7815

Qy      441 uAspArgLysValLeuPro---AlaMetLeuAspIleValArgArgArg----------- 456
               ||||||||||||         :::::::||||||::::::
Db     7816 AGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGAC 7875

Qy      457 ------------------------IlePro---------AlaAspLysLeuProAspIl 465
                                    :::|||         :::|||||||||
Db     7876 AGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTT-------- 7927

Qy      465 ePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPheVa 485
                        |||   :::|||        |||   |||||||||
Db     7928 ------------ATTGCTCGTAAAAAAGACTGGGATCCAAAAAAAATATGGTGGT----- 7969

Qy      485 lPheAspLysSerValValProTyrSerAspLysPheHisAla---------------- 499
              ||||||         |||   ||||||              |||
Db     7970 -TTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAA 8028

Qy      500 -------MetLeuLysSer-------------------------MetAspLysGluLy 508
                   |||||||||                           |||:::::
Db     8029 ATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTC 8088

Qy      508 sPheAla------------------------------------LysAlaIleGluLysAs 516
              |||                                      |||   :::::::|||||
Db     8089 CTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGA 8148

Qy      516 pProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAl 536
             |       ::::::|||   |||
Db     8149 CTTAATCATTAAACTACCTAAA-------------------------------------- 8170

Qy      536 aMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGl 556
                  :::    :::    :::|||   |||::::::          |||        ||
Db     8171 ----TATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGA 8226

Qy      556 uMetTyrProGlyArg------AlaLeuProSerAsp---AlaAsnPheThrMetArgMe 573
             |:::       |||           |||||||||||         ||||||
Db     8227 ATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGC 8286

Qy      573 tSer---TyrGlySerIleLysGlyTyrGluProGlnAspGly-----------AlaTr 588
             |||    |||       :::||||||       |||:::|||              ::
Db     8287 TAGTCATTATGAAAAGTTGAAGGGT---AGTCCAGAAGATAACGAACAAAAACAATTGTT 8343

Qy      588 pTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGl 608
             :            |||         |||         |||::::::           :::       ||
Db     8344 TGTGGAGCAGCAT---------AAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGA 8394

Qy      608 uPheAlaValGln----------------------------------------------- 612
             ||||:::   :::
Db     8395 ATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAA 8454

Qy      613 ---------------------------------GluAsnIleLeuAspLeuPheArgTh 621
                                             |||||||||:::   ||||||
Db     8455 CAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTT 8514

Qy      621 rLysAsnTyrGly---------------------------------------------Ar 626
                  |||   |||                                             ||
Db     8515 GACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTGATACAACAATTGATCGTAAACG 8574

Qy      626 gTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly 645
             ||||          ::::::      |||   |||    :::     ||||||||
Db     8575 ATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGT 8632

RESULT 6
AE000733
LOCUS       AE000733                15569 bp    DNA     linear   BCT 25-MAR-1998
DEFINITION  Aquifex aeolicus section 65 of 109 of the complete genome.
ACCESSION   AE000733 AE000657
VERSION     AE000733.1  GI:2983720
KEYWORDS    .
SOURCE      Aquifex aeolicus.
  ORGANISM  Aquifex aeolicus
            Bacteria; Aquificae; Aquificae; Aquificales; Aquificaceae; Aquifex.
REFERENCE   1  (bases 1 to 15569)
  AUTHORS   Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,
            Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M.,
            Huber,R., Feldman,R.A., Short,J.M., Olsen,G.J. and Swanson,R.V.
  TITLE     The complete genome of the hyperthermophilic bacterium Aquifex
            aeolicus
  JOURNAL   Nature 392 (6674), 353-358 (1998)
  MEDLINE   98196666
   PUBMED   9537320
REFERENCE   2  (bases 1 to 15569)
  AUTHORS   Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,
            Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M.,
            Huber,R., Feldman,R.A., Short,J.M., Olson,G.J. and Swanson,R.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego,
            CA 92121
COMMENT     Putative indicates no similarity to known proteins
            Hypothetical indicates similarity to a protein of unknown function.
FEATURES             Location/Qualifiers
     source          1..15569
                     /organism="Aquifex aeolicus"
                     /strain="VF5"
                     /db_xref="taxon:63363"
     gene            complement(199..1497)
                     /gene="purA"
                     /note="aq_1290"
     CDS             complement(199..1497)
                     /gene="purA"
                     /codon_start=1
                     /transl_table=11
                     /product="adenylosuccinate synthetase"
                     /protein_id="AAC07286.1"
                     /db_xref="GI:2983724"
                     /translation="MEKLIILGAQWGDEGKGKIVDLLSEHFDITVRYQGGSNAGHTVV
                     VNSQKFILHLLPTGILHEHVKGVIAQGMVVDLEVLHKEVKNLEEKGIYVKERLFISDR
                     AHLVMPYHKLLDSLFEKKKGIGTTLRGIGPAYMFKYGRKGIRISDLKDEKRFYTLLED
                     NLDFVKNICEKVFCEKFDLDINQIYEEQLRYFEEFKENVVDLLRFFNTQKGSVLFEGA
```

```
                QGTLLDVDMGTYPYVTSSNASALGLSNGTGMPPKYFSDAFFLGVAKAYTTRVGEGPFP
                TELKGEEGEKLRELGGEYGSTTGRPRRCGWLDLVALKYAVQVNGLDGFVITKLDVLDT
                FDEVKVCVAYELDGEVIDYFPASYSELIRVKPVYKTLKGWKKSTKGAKDVSELPKEAL
                DYVKFIEEYTGVPVVMLSTGPKRDEYIWLKEILRTRSGYS"
gene            complement(1526. .2062)
                /gene="aq_1292"
CDS             complement(1526. .2062)
                /gene="aq_1292"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAC07294.1"
                /db_xref="GI:2983732"
                /translation="MKIIITGEPGVGKTTLVKKIVERLGKRAIGFWTEEVRDPETKKR
                TGFRIITTEGKKKIFSSKFFTSKKLVGSYGVNVQYFEELAIPILERAYREAKKDRRKV
                IIIDEIGKMELFSKKFRDLVRQIMHDPNVNVVATIPIRDVHPLVKEIRRLPGAVLIEL
                TPENRDVILEDILSLLER"
gene            2131. .4734
                /gene="alaS"
                /note="aq_1293"
CDS             2131. .4734
                /gene="alaS"
                /codon_start=1
                /transl_table=11
                /product="alanyl-tRNA synthetase"
                /protein_id="AAC07289.1"
                /db_xref="GI:2983727"
                /translation="MSLSAHEIRELFLSFFEKKGHTRVKSAPLVPENDPTLLFVNAGM
                VPFKNVFLGLEKRPYKRATSCQKCLRVSGKHNDLEQVGYTSRHHTFFEMLGNFSFGDY
                FKKEAIEYAWEFVTEVLKLPKEKLYVSVYKDDEEAYRIWNEHIGIPSERIWRLGEEDN
                FWQMGDVGPCGPSSEIYVDRGEEYEGDERYLEIWNLVFMQYNRDENGVLTPLPHPNID
                TGMGLERIASVLQGKNSNFEIDIIFPLIQFGEEVSGKKYGEKFETDVALRVIADHLRA
                ITFAISDGVIPSNEGRGYVIRRILRRAMRFGYKLGIENPFLYKGVDLVVDIMKEPYPE
                LELSREFVKGIVKGEEKRFIKTLKAGMEYIQEVIQKALEEGRKTLSGKEVFTAYDTYG
                FPVDLIDEIAREKGLGIDLEGFQCELEEQRERARKHFKVEAKKVKPVYSHLKELGKTS
                AFVGYEHMEWESQVVGLVKGEGLVSELKEGEEGEVVLKETPFYPEGGGQIGDAGIIES
                DKALFKVEDTQKPTEGIIVHIGKVLKGTLKVGDTVHARVDKERRWDIMRNHTATHLLH
                AALRNVLGEHVRQAGSLVADKYLRFDFTHFSALTEEELKRVEELVNEKIRENLPVNVM
                EMAYDEALKTGAIAIFEEKYGERVRVISCGEFSKELCGGTHVSATGDIGYFKIISESS
                VGAGVRRIVAQTGRWSVETAFKEHQTLKKASSALGVGEEEVIQKIEELKEEIKDRERE
                IQRLKQELLKLQIREVVKEENVGDFTLHYGVFEEVEPEELRNLADMLRQRTKKDVVFI
                ASRKGDKINFVIGVSKEISDKVNAKEVIREVGKVLKGGGGGRADLAQGGGKAPDKFPE
                AVKLLKEILSG"
gene            4739. .5209
                /gene="aq_1295"
CDS             4739. .5209
                /gene="aq_1295"
                /codon_start=1
                /transl_table=11
                /product="putative protein"
                /protein_id="AAC07293.1"
                /db_xref="GI:2983731"
                /translation="MMKEETRTGIGLKDVSIRLVSESPLLPVGGFVYCEYGIPVRYLK
                IDENEVKLEISVEKLKNLSRVCENIEQVVDKVVEELRYALPEGVFTIAKVYANGIKLR
                ELSYLEEYTSVGKGIIIDEDRELDQEMPAYVKENRETILGTIREVILLELMYSS"
gene            5306. .7738
                /gene="clpC"
                /note="aq_1296"
CDS             5306. .7738
                /gene="clpC"
                /codon_start=1
                /transl_table=11
                /product="ATP-dependent Clp protease"
                /protein_id="AAC07290.1"
                /db_xref="GI:2983728"
                /translation="MFEKFTERARQVILRAREEALELGHSYLGSEHILLALIKDEDIP
                VLVLSKLGVSPEKIRKAIMGQITKGSHSGEVLFAPDAKRVIEFAVEEARILHHQFVGP
                EHLLIGVVREKTGLGGRILRSFGLDEYSVRREVLQLLGELPPQESQRYVSTPNLDRFS
                RDLTQMAREGKLDPVIGREKEIERVIQILVRRRKNNPVLLGDPGVGKTAIVEGLAQRI
                ANKEVPEPLLNKRVVALDLAALVAGTKYRGQFEERLKNILKELEKANNVILFIDELHT
                LVGAGSAEGSIDASNMLKPALARGEIQVIGATTVDEYRKYIEKDGALERRFQPVYVEQ
                PSPEDTIQILYGLKPKFEEFHKVEYTPEAIEKAVELSVKYITDRNLPDKAIDVMDEAG
                ALVKLKATQLPPELKEIEEKIKELEKEKEKAVKEQDYEKAAKIRDEELRLRAKFETMK
                AKWKEEMEKNRPKVTDEDVAEVVARWTGVPVKRVHESDAEKLLHIEEELKKRVVGQDE
                AIKAVAKAIRRSRVGLKGKHRPIGVFMFLGPTGVGKTETAKALAEYLFGTEDALIRFD
                MSEYMEKHTVSRLIGAPPGYVGYEEGGQLTEAVRRRPYSVLLFDEIEKAHPDVFNIFL
                QIFDDGRLTDGLGRTVDFSNTIIIMTSNLGARLITGGTMGFEKKQGLLNFEEMRKNVL
                DQVKRTFSPEFLNRLDEIIVFRPLEKKDIEQILELQLKEINEKLKDWNVRVKLHKSFV
                NWIIEKEYKPEFGARSLKRALQRHVEDLLAEELIKGTLKGVEEVEIRVRNDKPYLKPV
                KKKEKVEALLEK"
gene            7776. .10112
                /gene="omp"
                /note="aq_1300"
CDS             7776. .10112
                /gene="omp"
                /codon_start=1
                /transl_table=11
                /product="outer membrane protein"
                /protein_id="AAC07292.1"
                /db_xref="GI:2983730"
                /translation="MKKEFLTLALFLGVANAAEIEKIIIKGNKYIPDELIRGLLSVRK
                GSEFSLERIRKDIRKLYRTGLFTKVEAYEEDQNKNGKPEVILVVEDLPVIYKIEFEGN
                EEFSDDDLKEYLGIETELGEVEVEETIEGYTSSPAIEERLAVMKQLKLGRILTFQEIE
                SLRRRIEEVYKRHGFVGTKVEYKLIPKKGASKLVFVIHEGKKIYTSDIEIEGNKHFSD
                RKLKGLMELKEPNIFLFRLHPAFSEETLKEDVKKIEEFYKNEGFLEVEVSYEVKEEGP
                ARKVIVKVKEGPRYKLKEIKIEGNTLFAYSELVDNILKKNERKGRYFRREVIEKIKNR
                IREKYAEIGFLNTSVEERVNVNPEKKEVSVLLKIIEGKPVYVKKIKIKGNYETRDYVI
                RREMRVQENELALKKGIERSKTRIMNLGYYEDVQIEPVPRRDAWWDLLVKIRERFTGQ
                FSVGLSYNEVTGLAGFIELRKGNFRGTGDIAGISVSYGSLYRNNAISYTRKWFLKKPV
                DLDLSAFDRRIEYDTYTVERTGFSVALSKELSEYWRASIGTSIQRVKYSDIDPEASTY
                VKEQAGRRDSRKIFFTLTRDTRDYYLLPTKGSLFVFRNSVGVGLLGGDEKFYKFEVEG
                AKYFSDTYFDTGIILSLKGEVGFVEGYGGKKVPIDERFFVGGDFSIRGYEYGYAGPVD
                PNTLDPIGAKKKVITSLELMYPVFKKMLYFAGFFDYGLGADKWSDFKLSNFRGGYGIG
                VRIITPFAPIRIDWAFKTKTVPGDTNRSRIHFVLGTFF"
gene            complement(10138. .10350)
                /gene="cspC"
                /note="aq_1303a"
CDS             complement(10138. .10350)
                /gene="cspC"
                /codon_start=1
                /transl_table=11
                /product="cold shock protein"
                /protein_id="AAC07291.1"
                /db_xref="GI:2983729"
                /translation="MSFRGTVKWFSKDKGYGFITREDTNADVFVHFTDIQMEGFKTLQ
                KGQKVEFDVVEDTKGPRAKNVRVLGE"
gene            complement(10454. .11170)
                /gene="hisA"
                /note="aq_1303"
CDS             complement(10454. .11170)
                /gene="hisA"
                /codon_start=1
                /transl_table=11
                /product="phosphoribosylformimino-5-aminoimidazole
                carboxamide ribotide isomerase"
                /protein_id="AAC07284.1"
                /db_xref="GI:2983722"
                /translation="MNLKEFIIPAIDLMEGKAVRLYKGDFNKVKVYSERPWELAKNFS
                DLGFKRLHVVDLEGAEGGKPKNLEVIRKIRENFEGEVEVGGGIRSYEVAKALFDEGID
                FVVIGTLAYKNKEEFLKILENFPNRVILAIDSKQGKVAIGGWKEETAVSPEEFAKEYE
                NYPIWGYLYTVIERDGSLEGVDVEPYKEIKKHVKKPVIASGGVSSLEDIKKLYGIVEG
                VVVGKAIYEGRITLEDLQNF"
gene            complement(11167. .11922)
                /gene="pyrDB"
                /note="aq_1305"
CDS             complement(11167. .11922)
                /gene="pyrDB"
                /codon_start=1
                /transl_table=11
                /product="dihydroorotate dehydrogenase electron transfer
                subunit"
                /protein_id="AAC07287.1"
                /db_xref="GI:2983725"
                /translation="MLAELNAEVLENTYISGNLYRLTLRIPDKILKQIEPGHFAMIKP
                SDTYDPMGRRAFAVADIEEHKAVFYYDVYGKGTYLLSKRKPGERLKTFLPLGKRLFSY
                EGDKHLLIGGGVGLAGLTLLAKKLRDMGKKVFIAYGGRSREHLGMVEWLEKEEFPYEL
```

```
                FTNDGSAGRKGFVTEILKEFSTDWVVHACGPKPMLKTIKEMKTGHKVYFSLEERMACG
                WGVCLGCVVKTRDGKFKRVCYEGPVMPMEEVIL"
     gene       11990..13300
                /gene="sucCl"
                /note="aq_1306"
     CDS        11990..13300
                /gene="sucCl"
                /codon_start=1
                /transl_table=11

Alignment Scores:
Pred. No.:              4.82         Length:          15569
Score:                  144.50       Matches:         170
Percent Similarity:     34.18%       Conservative:    111
Best Local Similarity:  20.68%       Mismatches:      267
Query Match:            3.89%        Indels:          277
DB:                     1            Gaps:            40

US-10-008-355-2 (1-712) x AE000733 (1-15569)

Qy        9 LeuLeuGlyAlaAlaLeu------LeuLeuGly-----------AlaSerGlyValAla 22
             ||||||  |||||||||      ::||||||            :::|||
Db     3820 CTGCTTCACGCCGCACTCAGGAACGTTCTCGGCGAGCACGTAAGACAGGCTGGTTCTCTC 3879

Qy       23 LysAlaAspLysGlyMetTrpLeu-----------LeuAsnGluLeuAsnGlnGluAsn 38
               |||||||||   :::                  :::   |||   :::|||
Db     3880 GTCGCGGACAAGTATTTGAGGTTTGACTTTACACACTTTTCAGCTCTAACGGAAGAGGAA 3939

Qy       39 LeuAspArgMetArgGluLeu--------------GlyPheThrLeuProLeuAspSer 53
             |||   |||:::   ||||||                          ||||||:::::
Db     3940 CTCAAGAGGGTGGAAGAACTCGTGAACGAAAAAATAAGGGAAAACCTCCCCGTGAATGTA 3999

Qy       54 Leu---TyrSerPheAspLysProSerIleAlaAsnAlaValValIlePheGlyGlyGly 72
            :::      ::::::|||:::                  |||:::   ||||||
Db     4000 ATGGAAATGGCATACGATGAAGCACTAAAAACGGGTGCCATAGCGATATTTGAGGAAAAG 4059

Qy       73 ------------------------------------------------CysThrGly 75
                                                            |||  |||
Db     4060 TACGGGGAAAGGGTGAGGGTAATATCCTGCGGAGAGTTTTCAAAGGAACTCTGCGGTGGA 4119

Qy       76 IleThrValSerAspGlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIle 95
                  ||||||       |||   |||                ||||||     |||
Db     4120 ACGCACGTTTCCGCAACTGGAGACATA-----------------GGCTACTTCAAGATA 4161

Qy       96 GlnSerGlnSerThrVal----------------------------------------- 101
               |||:::|||:::|||
Db     4162 ATATCCGAGAGCTCCGTGGGTGCGGGCGTGAGGAGAATAGTGGCACAAACGGGAAGGTGG 4221

Qy      102 --------------------AspHisAspTyrLeuArgAspGlyPheValSerArgThr 114
                                :::|||      |||:::               :::
Db     4222 TCCGTAGAAACCGCCTTCAAAGAACACCAGACGCTCAAGAAAGCTTCTTCAGCTCTCGGC 4281

Qy      115 MetGlyGluGluLeuProIleProGlyLeuSerValLysTyrLeuArgLysIleValLys 134
            :::|||||||||                                    ::::::::::::
Db     4282 GTGGGCGAGGAG-----------------------------------GAAGTAATCCAG 4305

Qy      135 ValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArg 154
               ::::::::::::   ::::::|||           |||::::::||||||:::
Db     4306 AAAATTGAAGAACTTAAGGAAGAAATCAAAGACAGAGAAAGAGAAATACAAAGACTGAAA 4365

Qy      155 LysAlaGlnGluVal-----------CysGlnGluLeuAlaLysLysGluAsnAlaAsp 170
               ||||||:::                 :::|||:::   |||:::||||||
Db     4366 ------CAGGAACTCCTGAAACTCCAGATAAGGGAAGTTGTAAAGGAAGAAAACGTTGGA 4419

Qy      171 GluAsnGlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyrPheLeuIleVal 190
            :::                                             |||   :::
Db     4420 GAT-------------------------------------------TTCACACTCCAC 4434

Qy      191 TyrAspValPheLysAspValArgMetValPheAlaProProSerSerValGlyLysPhe 210
            |||   ||||||::::::|||                  |||         :::
Db     4435 TACGGCGTATTTGAAGAGGTTGAA---------------CCCGAAGAACTCAGGAACCTC 4479

Qy      211 GlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArg 230

Db     4479 ------------------------------------------------------------ 4479

Qy      231 ValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLys 250
                           ||||||                        ||||||
Db     4480 -----------GCGGATATGCTGAGGCAAAGGACGAAAAAGGAC-------------- 4512

Qy      251 ProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThr 270
                                |||   :::         |||  |||
Db     4513 --------------GTTGTGTTTATAGCCTCCAGAAAGGGAGATAAGATAAACTTCGTT 4557

Qy      271 IleGly-PhePro------------------------------------GlySerTh 277
            |||||| ||||||                                     |||   ::
Db     4558 ATCGGAGTTTCCAAGGAGATATCGGACAAGGTGAACGCCAAGGAAGTAATAAGGGAGGTC 4617

Qy      277 rAspArgTyrLeu-----------------------ThrSerTrpGlyValGluAspAr 289
             :::::|||||||||                      ||||||    ||||||:::||
Db     4618 GGAAAGGTACTTAAAGGTGGAGGAGGAGGAAGGGCAGACCTCGCTCAGGGTGGAGGAAAG 4677

Qy      289 gIleGluAsnGluAsnAsnPro-----------------ArgIleGluValArgGlyIl 303
             |:::         |||:::|||                 |||        ||||||
Db     4678 GCTCCTGACA---AATTCCCCGAAGCGGTTAAATTACTAAAGGAAATCTTATCGGGGT-- 4732

Qy      303 eLysGlnGlyIleTrpLysGluAlaMetSerAla-----------------AspGlnAlaTh 318
               :::|||:::  ||||||                          |||   :::
Db     4733 ----AAAGGTGTGATGAAGGAAGAGACACGAACGGGTATCGGACTCAAAGACGTGAGTAT 4788

Qy      318 rArgIleLysTyrAlaSer-------------------------------LysTyrAl 327
             |||:::      |||                                 :::|||
Db     4789 AAGGCTCGTGAGCGAGTCTCCCCTCCTTCCCGTGGGGGGCTTCGTTTACTGTGAGTACGG 4848

Qy      327 aGlnSerAlaAsnTyrTrpLys---------AsnSerIleGlyMetAsnArgGlyLeuAl 344
                         |||  |||           |||   :::   :::         :::
Db     4849 AATCCCCGTAAGGTATCTGAAGATAGACGAAAACGAAGTAAAACTCGAGATTTCCGTGGA 4908

Qy      344 aArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleAr 364
             :::|||      :::   |||                            |||  :::
Db     4909 AAAGCTGAAAAATTTATCGAGGGTGTGTGAAAACATAGAGCAAGTTGTTGACAAAGTTGT 4968

Qy      364 gLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLy 384
             :::      :::   |||:::        |||         ::::::   |||   |||
Db     4969 AGAAGAGTTGAGGTACGCCCTGCCCGAGGGTGTTTTC---ACTATAGCGAAGGTTTACGC 5025

Qy      384 sGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyTh 404
                |||   |||      ||||||::::::||||||   |||
Db     5026 AAACGGTATTAAACTC---AGAGAACTCTCTTACCTTGAAGAG---------------- 5065

Qy      404 rGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAl 424
                          :::                 :::       :::   |||
Db     5066 -------------TACACTTCCGTGGGTAAAGGAATAATAATAGACGAGGAC-------- 5104

Qy      424 aGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLy 444
                      :::   ||||||   :::            |||:::
Db     5105 ---------AGGGAACTGGATCAGGAAATGCCCGCTTACGTTAAGGAAAACAGGGAAAC 5154

Qy      444 sValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeuProAs 464
             :::|||      :::   :::::::::            :::                ||
Db     5155 GATCCTCGGCACTATCAGGGAAGTCATCCTGCTTGAGCTCATGTACTCTTCTTAAAAAGA 5214

Qy      464 pIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPh 484
             |||   |||   :::   :::         |||      ||||||            ||
Db     5215 CATTATAAAG---ATATATAACACTGTTTTTGTCATAGCCTTGAAAAAGCTTAGGTTTTT 5271

Qy      484 eValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLysSerMe 504
            |:::|||   |||      :::                |||   |||
Db     5272 TTTATTTTATAAAAAACTGAGAGAGGTAAGAGACATGTTT------------------- 5311

Qy      504 tAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLysSe 524
                  |||||||||                                |||   :::::
Db     5312 -------GAGAAGTTC--------------------------ACAGAAAGGGCAAGACA 5337

Qy      524 rValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGl 544
             ||||||       ||||||                ::::|||:::          ::::::
```

```
Db    5338 AGTTATCTTGAGGGCGAGG---------GAAGAAGCCCTCGAGCTGGGACACAGCTACTT 5388

Qy     544 uLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuPr 564
                 :::   :::       |||   ::::::::          |||       ||
Db    5389 GGGAAGTGAACATATTTGCTTGCTCTGATAAAGGATGAGGA-CATACCCGTCCTCGTCC 5447

Qy     564 oSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGlu---Pr 583
           |     |||                :::|||          ||||||:::       ||
Db    5448 TCAGCAAGCTGGGTGT---------CTCTCCTGAGAAGATAAGAAAGGCCATAATGGGCC 5498

Qy     583 oGlnAsp---------------GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGly-- 597
           |  :::                |||      |||:::   |||   |||   ||||||
Db    5499 AAATAACTAAGGGAAGCCACTCGGGAGAGG-TACTCTTCGCACCCGACGCAAAAAGGGTA 5557

Qy     597 ------------------------------------------------------------ 597

Db    5558 ATAGAGTTTGCAGTTGAAGAAGCGAGAATTCTTCACCACCAGTTCGTAGGTCCCGAACAC 5617

Qy     598 ------------ValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGl 613
                           |||||||||:::              :::||||||
Db    5618 CTACTTATCGGAGTGGTTAGAGAAAAGACGGGACTCGGTGGAAGAATTT---------TA 5668

Qy     613 uAsnIleLeuAspLeuPheArgThr---------LysAsnTyrGlyArgTyrAlaGluAs 630
                 ||||||         |||         :::   |||       |||:::|||||
Db    5669 AGGAGCTTTGGACTGGACGAGTACTCCGTAAGGAGAGAGGTACTTCAACTACTCGGAGAA 5728

Qy     630 nGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySe 650
           |     |||     |||                         |||          ::
Db    5729 CTTCCCCCTCAAGAAAGCC-----------------------AGAGGTACGTA-TCCAC 5763

Qy     650 rProValPheAspLysAsnGlyArg---LeuIleGlyLeuAlaPheAspGlyAsnTrpGl 669
           :|||      |||:::      |||   |||       :::|||   :::|||       ::
Db    5764 TCCTAACCTTGACAGGTTCTCAAGGGACCTCACCCAAATGGCAAGGAAGGAAAGCTCGA 5823

Qy     669 uAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerValAspIl 689
           :   :::   |||                |||   :::::::::|||   |||   :::   ::
Db    5824 CCCAGTAATAGGCAGG---------GAAAAGGAAATAGAGAGAGTTATCCAGATACTCGT 5874

Qy     689 eArg 690
           :|||
Db    5875 AAGA 5878
```

RESULT 7
OSJN00169/c

```
LOCUS       OSJN00169          162754 bp    DNA     linear   PLN 17-JUL-2002
DEFINITION  Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBa0043L24,
            complete sequence.
ACCESSION   AL662969
VERSION     AL662969.2  GI:21912514
KEYWORDS    HTG.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q., Hu,X.,
            Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J., Zhang,L.,
            Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.J., Lu,Y., Li,C.,
            Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K., Zhou,B.,
            Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P., Fu,G.,
            Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu,W.Q., Zhu,G.F., Tu,Y.F.,
            Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H., Chen,X.Y.,
            Shao,C.Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W., Wang,L.J.,
            Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L., Zhu,F.H. and
            Hong,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2002) Han Bin, National Center for Gene Research,
            Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
            CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
            bhan@ncgr.ac.cn
  REMARK    Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
            clone: OSJNBa0043L24.
COMMENT     On Jul 19, 2002 this sequence version replaced gi:17998481.
            Web site: http://www.ncgr.ac.cn
            -------------- Summary Statistics
            Assembly program: phrap
            --------------
            NOTE: This is a complete sequence.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Fgenesh (http://www.softberry.com/),
            genscan (http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM
            (http://genemark.biology.gatech.edu/GeneMark/), tRNAscan-SE (Sean
            Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/), searches of the
            complete sequence against NCBI none redundant protein database (nr)
            (ftp://ncbi.nlm.nih.gov/blast/db) and the EST database at NCGR.
FEATURES             Location/Qualifiers
     source          1..162754
                     /organism="Oryza sativa"
                     /variety="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:4530"
                     /chromosome="4"
                     /clone="OSJNBa0043L24"
                     /clone_lib="CUGI-OSJNBa"
BASE COUNT    45645 a  36513 c  35782 g  44812 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:              256            Length:          162754
Score:                  137.50         Matches:         147
Percent Similarity:     36.64%         Conservative:    115
Best Local Similarity:  20.56%         Mismatches:      260
Query Match:            3.70%          Indels:          195
DB:                     8              Gaps:            34

US-10-008-355-2 (1-712) x OSJN00169 (1-162754)

Qy      47 PheThrLeuProLeuAspSerLeuTyrSerPheAspLysProSerIleAlaAsnAlaVal 66
           |||      |||      ||||||                  |||   |||:::
Db   53593 TTTGTGCAGCCTTCTCGTTCTTTA-----------------TCACATGCAAGTGGTACA 53552

Qy      67 ValIlePheGlyGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThr 86
                 |||   |||   :::   |||         |||:::
Db   53551 AGTGCACTTGGATATGGCACTTCAGCTATTGTTGCTATGGACAAA--------------- 53507

Qy      87 AsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThrValAspHisAspTyrLeu 106
                             ||||||   |||   |||:::               :::   |||
Db   53506 ---------------ACTGCTGCTATTTCTTCAGATTCATCCTTTTCATCAAATACTCTT 53462

Qy     107 ArgAspGlyPheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSerVal 126
           |||               |||:::|||:::|||                           |||
Db   53461 CGTTTGTCA---CAGTCAAAGACTGTTGGGAGAAGT-----------------TCTGAA 53423

Qy     127 LysTyrLeuArgLysIleValLysValThrAspLys-----------ValGluGlyGln 142
           :::   |||      ::::::      :::   :::             :::|||
Db   53422 AGAAGCCTAGAGAGTGTGCTTAACTCAAGCAAAGAAAAGGTTTCTGCCATTGAAAGTTTG 53363

Qy     143 LeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGlu 162
           |||||||||:::::
Db   53362 CTGAAAGGTGTAAGC--------------------------------------------- 53348

Qy     163 LeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyrSer 182
           ::::::   ::::::|||                :::         ::::::   :::||||||
Db   53347 ATATCAGATAGGCAAAATATCTCAGCTACCCGCTCAACCAGCTTGGATCT-TGGTATTCT 53289

Qy     183 AsnAsn-----------GluTyrPheLeuIleVal-----------TyrAspValPhe 194
              |||            |||      ::::::             |||   ||||||
Db   53288 TTTAACTGCTTAGTTGTCAATTACTGTTAAGTAATAGAAATGGCCTTTTATCATGTATTT 53229

Qy     195 LysAsp------ValArgMetValPheAlaProProSerSerValGlyLysPheGlyGly 212
                       :::   ::::::   :::         |||   :::   |||
Db   53228 CCTGGTCATTGACTGACAATATTATCATCTTAGTTTTCTATGATAAAAAAAATC------ 53175

Qy     213 AspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyr 232
                 ::::::::::::|||                                     |||
```

```
Db   53174 ---------AGTTTCTTATTTCCT-----------------------TTTAATCCTAGT 53148

Qy     233 AlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProVal 252
           ||||||   |||                        |||:::|||          ||||||
Db   53147 GCAGGGGTTGAT---------CCTCCATCGTCTCGTGAT------------CCTCCTGTT 53109

Qy     253 TyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGly 272
                 ||||||   :::                         :::   |||:::
Db   53108 CCGCTTGCAGCAACAGCTTCAAATCATCTGTCCTTACAGAATTCAGCACTGCTCGACTCA 53049

Qy     273 PheProGlySerThr---------------------------------------AspArg 279
                 ||||||                                         :::
Db   53048 TCTGTTCCTAGCACCATAAATGCTAGTGCACGAAATGGAGGCTCCCGCTTATTGGAGTCA 52989

Qy     280 TyrLeuThrSerTrpGlyValGluAspArg-----------------------IleGlu 291
              |||      |||       :::|||                              |||
Db   52988 ATGACAACCCAGTTGGGCACCAGAGAGCGATCAAGATCACCGTACTTGGGTAATATATCA 52929

Qy     292 AsnGluAsnAsnProArgIleGluVal------------ArgGlyIleLysGlnGlyIle 307
           :::|||:::          :::   :::              |||   :::::::::
Db   52928 TCTGAGTCCATGACAAGCTTATCACTGCCTTTTCCAAGAAGGTCTTTGGAGAGGCCA--- 52872

Qy     308 TrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAla 327
              :::|||            |||:::   :::   |||:::         ::::::
Db   52871 ---CAAGAAGGAGGCCGCATGGATGAGGGTAGTGATATCCGGTCAACTAGGAGATTCCCC 52815

Qy     328 GlnSerAlaAsnTyr------------TrpLysAsnSerIle------------------ 337
           |||:::   ||||||                 ::::::::::::|||
Db   52814 CAAACTCAAAACTATGTTGACATGCCCTACAGAGATGCCATCCACAGAGATTCACATAAT 52755

Qy     338 ------------GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArg 353
                          |||   |||   |||         |||:::|||   |||
Db   52754 AATCATGTTCCAAATTTCCAGAGACCTCTTCTAAGGAAGCAAGTAATGTCAAGGGCTTCT 52695

Qy     354 AlaGluGluArg---AlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyr 372
           |||         |||   :::|||   |||         :::::::|||      :::   |||
Db   52694 GCCAGCATCAGGCACAGCTTTGACGATAGCCAAGTACAGTCAGGTGACGTGTCTGGTTAT 52635

Qy     373 GlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGlu 392
              |||   |||:::||||||         |||      ||||||
Db   52634 ACGGATGCACTGGCTTCACTAAGTGATGCTCTTTCTGAGGGT------------------ 52593

Qy     393 MetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPhe 412
                    ||||||   :::                |||:::|||   :::   |||   :::
Db   52592 ---------CTCAGCCCTAGTTCGGATTG-GGTAGTGAGAGTCTCAGCTTTTGAATTTAT 52543

Qy     413 AlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGly-IleLeuLysSerLeuAspAs 432
           :::   :::   ||||||   |||:::   :::::::||| |||   ::::::   :::
Db   52542 TCGGAATCTTCTGCAACAAGGCCAGAGAGGTATTCAGGAAATTACACAGAATTTTGAAAA 52483

Qy     432 pLysTyrLysAspTyrLeuProSerLeuAsp---------ArgLysValLeuProAlaMe 449
               |||   :::         ||||||              ||||||      |||
Db   52482 GGTCATGAAGCTGTTTTTTCGTCATTTGGATGACCCTCATCATAAGGTTGCACAAGCAGC 52423

Qy     449 tLeuAspIleValArgArgArgIleProAlaAspLys---------------------- 461
                   :::            |||||||||   |||
Db   52422 CTTTTCTACTCTTGCAGAACTTATTCCAGCTTGCAAGAAGCCATTTGAGAGTTATGTTGA 52363

Qy     462 -------LeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLy 479
                  ||||||   :::|||      :::||||||   |||               ||
Db   52362 AAGAATTTTGCCATATGTTTTTTCCCGGCTTATTGATCCAAAAGAATTG------GTAAA 52309

Qy     479 sLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAspLys------Ph 497
            ||||                   |||            |||:::
Db   52308 AAAACCATGCTCATCAACCTTGGATGTTGTTGGGAGAACATATGCAATTGATATGTTGCT 52249

Qy     497 eHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspPr 517
               |||:::::::::|||:::|||:::::::::          ||||||      |||
Db   52248 ACCTGCCCTAGTGCGCTCACTAGATGAACAGAGGTCCCCAAAGGCA------AAGTTGGC 52195

Qy     517 oAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMe 537
                :::|||   :::                  :::         :::   ::::::
Db   52194 TGTTCTTGAGTTCGCAAATAAATCGTTC---AGCAAGTACACTGTAGATTCTGAAGGTTA 52138

Qy     537 tAlaAsnAlaTyrAlaIleGlu-------------------------------------- 544
             :::|||:::      ::::::
Db   52137 TAGTAACAGTGGGTTCCTTAAACTGTGGCTTTCAAAATTGGCACCTTTAGTTCATGAAAA 52078

Qy     545 -------LysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAl 562
                  |||   |||               :::|||:::      :::|||
Db   52077 AAATGCAAAACTGAAGGAGGCATCCATTTCTGGTATCATATCTGTTTATTCTCATTTTGA 52018

Qy     562 aLeuProSerAspAlaAsnPheThrMetArgMetSerTyr-------------------- 575
                :::        ||||||   :::   :::|||
Db   52017 TTCAACAGCAGTGCTAAATTTTATTCTAAACTTATCTGTTGAGGAGCAAAACCTCTTGAG 51958

Qy     576 -GlySerIleLysGlyTyrGluProGln---AspGlyAlaTrpTyrAsnTyrHisThrTh 594
                ::::::|||   |||   |||:::   :::               ||||||      ::
Db   51957 GCGTGCACTGAAGCAATACACACCTCGCATTGAAGTTGATTTAGTGAACTACTTACAGAG 51898

Qy     594 rGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAs 614
            :   |||         ::::::         |||||||||   :::   |||
Db   51897 CAAGAAA---------GATCGTCCCCGTCCCAAATCTTATGATCAAGCA----------- 51858

Qy     614 nIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHi 634
                                   :::||||||      :::|||:::|||
Db   51857 ------------------------GATTATGGAACTTCTTCCGAAGATGGT-------- 51831

Qy     634 sIleAlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlyProValPheAs 654
               |||      |||                     |||      |||         ||
Db   51830 -TACGCGCTTGCATCAAAGAAAAGCTATCCATTTGGGCGGTATTCCTCTAGTTCCCTTGA 51772

Qy     654 pLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrp 668
            |      |||                      |||   |||
Db   51771 TGCTGAAGGT------------------------GGGAAATGG 51753

RESULT 8
MSP207AAM
LOCUS       MSP207AAM               4004 bp    DNA     circular BCT 26-APR-1993
DEFINITION  M.sp 207 alpha-amylase gene.
ACCESSION   X55799
VERSION     X55799.1  GI:296762
KEYWORDS    alpha-amylase.
SOURCE      Micrococcus sp.
  ORGANISM  Micrococcus sp.
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Micrococcineae; Micrococcaceae; Micrococcus.
REFERENCE   1  (bases 1 to 4004)
  AUTHORS   Kimura,T. and Horikoshi,K.
  TITLE     The nucleotide sequence of an alpha-amylase gene from an
            alkalopsychrotrophic Micrococcus sp
  JOURNAL   FEMS Microbiol. Lett. 71, 35-42 (1990)
FEATURES             Location/Qualifiers
     source          1..4004
                     /organism="Micrococcus sp."
                     /strain="207"
                     /db_xref="taxon:1271"
     repeat_unit     165..189
                     /rpt_type=INVERTED
     RBS             265..272
     CDS             282..3596
                     /codon_start=1
                     /transl_table=11
                     /product="alpha-amylase"
                     /protein_id="CAA39321.1"
                     /db_xref="GI:296763"
                     /db_xref="SPTREMBL:Q06812"
                     /translation="MSKKWKQKFLMVLALLLVIGGDAVCSSATGQAETSTAPADREVR
                     LSILVRINDRDWIFGRGHRRRLMIKSILSRMKVGVRVSRHVAVAPRNEGIWICFAQHL
                     KLDYSPKRIWGSLIQVNKDSLTQAFVTSGIEQIRIVPDGSAPVIDNGENAMFYYRDKE
                     LFAADAMNTIEKVELQFDGETHEMTYEPENERFVFLANDIPTGEHPYTYLVTKAGVTT
                     EVTDPYNTVDGVSTLVFNQADLAVSGTVAPAAIDFNQNAVLSVDIEGITEDVKISKIV
                     ADTSLLGGSAQLEIDPTLKEISIAVDDDIAAGTKAIPLTVTDSYGNLYKGTAEVEVKT
                     RQSVGEGDFDWDESVIYFMLTDRFFDGDASNNDPYDLNYDKESRGTYQGGDFKGITDK
                     LDYLDELGVNTIWISPVVENIKYDVRYYETSEPYYGYHGYWANNFGELNPHFGTMEEF
                     HDLIDGAHDRNMKIMVDVVVNHTGYGLKEIDGSVTNPPAGYPSDADRARFSDLLRQGA
```

```
               DVGTDEVVGELAGLPDFITEDPNVRKQIIDWQTDWIEKATTENGNTIDYFRVDTVKHV
               EDATWMQFKNALTEKMPEFKMIGEAWGAKVDNTLGYLETGTMDSLLDFGFKETARSFV
               NGSLEAANASLTARNAKLDNTATLGQFLGSHDEEGFLHSLAGDKGKLQVAATLQATAK
               GQPVIYYGEELGQTGANNYPQYDNRYDFAWDQVEGNEILAHYTKILNFREGYSKVFAK
               GERTLVGGSDKDQFLLFSRDYQDQKVYVGLNVAEESKAVTLTVDSADAVVTDAYSGTE
               YTATAGKVNLTLPGKADPGTVLLTVEGGNITGVAKDNGEVVVELVPENNIRIHYKRED
               NVYKNYGAWLWNDVASPSANWPVGATMFEKTDSYGAYIDVPLADGAKNIGFLVMDITA
               GDAGKDGGDKGFTISSPQANEIWIKQGSDKVYTYEPVDLPANTVRIHYTREAVDYDDF
               GIWNWGDVASPSDGWPTGAVNFDGTDRYGAYPDIKLTEAAKEIGFLVVNEKTQEKDAG
               DKSFNLLDKFNHLWIKQGDDTVYVSPYDEVASGLGSAEVVSENTLLLRFNSTDGFTSE
               KLVAGLAIQDADSAAL"
BASE COUNT     1239 a    724 c    1028 g    1013 t
ORIGIN

Alignment Scores:
Pred. No.:              5.35          Length:          4004
Score:                  133.50        Matches:         138
Percent Similarity:     34.94%        Conservative:    94
Best Local Similarity:  20.78%        Mismatches:      259
Query Match:            3.59%         Indels:          173
DB:                     1             Gaps:            35

US-10-008-355-2 (1-712) x MSP207AAM (1-4004)

Qy     141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160
           |||      ||||||||||||||:::::::::     |||
Db    1431 GGCGATTTTAAAGGCATTACAGACAAGCTGGATTATTTA-------------------- 1469

Qy     161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180
                                    :::            |||   :::   |||      |||
Db    1470 -----------------GATGAGCTTGGGGTTAATACAATTTGGATT---AGTCCGGTT 1508

Qy     181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200
                       |||    :::|||                           |||||||||
Db    1509 GTGGAAAACATTAAGTAC--------------------------GATGTTCGTTATTAC 1541

Qy     201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrp------Met 218
                                 |||         |||           :::|||:::         :::
Db    1542 GAAACGTCAGAACCGTATTATGGCTATCACGGTTACTGGGCAAATAATTTTGGAGAGTTA 1601

Qy     219 TrpProArgHisThrGlyAspPheSerValPheArg---ValTyrAlaGlyAlaAspAsn 237
              |||   |||    |||               |||       :::     |||||    :::
Db    1602 AATCCG---CATTTTGGCACGATGGAAGAGTTTCATGACTTGATCGACGGGGCACACGAC 1658

Qy     238 ArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaVal 257
           |||       :::         |||                                  |||
Db    1659 CGCAACATGAAAATTATGGTCGAT--------------------------GTGGTGGTC 1691

Qy     258 SerMetGlnGlyTyr------LysAlaAspAspTyrAlaMetThr---------IleGly 272
           :::      ||||||         |||      |||   ::::::|||              |||
Db    1692 AATCACACAGGATACGGCTTAAAAGAAATTGATGGTTCGGTCACTAATCCGCCAGCGGGC 1751

Qy     273 PheProGlySerThrAspArgTyrLeuThrSer--------------------------- 283
           :::|||          ||||||         |||
Db    1752 TATCCGTCAGATGCCGACCGGGCAAGGTTTAGTGATTTGCTTCGTCAAGGGGCGGACGTA 1811

Qy     284 ---------------------------TrpGlyValGluAspArgIleGluAsnGluAsn 294
                                         |||:::   |||   |||      :::
Db    1812 GGTACAGATGAAGTGGTTGGTGAATTAGCGGGACTTCCAGACTTTATCACGGAAGATCCA 1871

Qy     295 AsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAla 314
           |||   |||   ::::::               |||      |||   |||      :::
Db    1872 AACGTTCGCAAACAAATTATCGAT---TGGCAAACCGATTGGATTGAAAAAGCGACGACA 1928

Qy     315 AspGlnAlaThrArgIleLysTyr------------AlaSerLysTyrAlaGlnSerAla 330
           :::               |||   |||                  |||:::   :::   |||
Db    1929 GAAAACGGCAATACGATTGATTATTTCCGTGTGGATACGGTGAAACACGTAGAAGATGCG 1988

Qy     331 AsnTyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGly 350
              :::   :::                 :::             :::           :::||||||
Db    1989 ACGTGGATGCAGTTTAAAAATGCATTAACGGAAAAAATGCCGGAATTTAAAATGATCGGT 2048

Qy     351 ArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAla 370
                                                          |||   |||      :::
Db    2049 GAAGCT--------------------------TGGGGCGCTAAAGTGGACAATACTTTA 2081

Qy     371 ValTyr---------GlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAla 387
              |||                 |||      :::       |||   |||:::         :::||||||   |||
Db    2082 GGCTATCTTGAAACGGGCACGATGGATTCATTGCTAGACTTTGGTTTTAAAGAAACGGCT 2141

Qy     388 Lys--------------------------AlaAsnArgGluMetThrTyrLeuSerGlu 398
           :::                              ||||||      :::|||          :::
Db    2142 CGGTCATTTGTTAACGGCAGTTTAGAAGCGGCTAATGCGTCTCTTACTGCACGAAACGCG 2201

Qy     399 ThrLeuPheGlyGlyThrGluValValArgPhe-----------------AlaGlnPhe 412
              |||                 :::   :::|||                                 |||
Db    2202 AAATTGGATAATACGGCAACACTCGGGCAGTTTTTAGGAAGCCATGACGAAGAAGGGTTC 2261

Qy     413 AlaAsnAlaLeuAlaThrAsn--------------------------------------- 419
              ::::::||||||   :::
Db    2262 CTCCATTCACTTGCTGGGGATAAAGGAAAGTTACAAGTCGCAGCCACTCTTCAAGCAACA 2321

Qy     420 -----------ProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLys 435
                         |||      :::   |||         :::   |||
Db    2322 GCAAAAGGACAACCTGTGATTTACTACGGA------GAAGAGCTTGGCCAAACAGGAGCT 2375

Qy     436 AspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArg 455
           :::      |||      |||   :::                                 |||
Db    2376 AATAATTATCCACAATACGATAATCGT--------------TATGATTTTGCGTGGGAT 2420

Qy     456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
           ::::::      :::::::::                  :::      ::::::::   :::   :::
Db    2421 CAAGTGGAAGGCAATGAAATTTTGGCGCATTACACCAAAATTTTAAACTTTAGA---GAA 2477

Qy     476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
           |||   :::|||   :::|||               ::::::::::::|||      ||||||
Db    2478 GGTTATTCGAAAGTGTTTGCGAAA------GGTGAACGCACGCTTGTTGGAGGATCTGAT 2531

Qy     496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
           |||         :::|||   |||   |||   :::                        :::
Db    2532 AAAGATCAATTCCTCTTATTTTCTCGTGACTATCAAGATCAAAAAGTGTACGTTGGGTTA 2591

Qy     516 AspProAlaValGluLeuSerLysSerVal---------IleAlaAlaAlaArgAlaIle 532
           :::   |||   |||   ||||||:::|||                    :::   :::|||   |||:::
Db    2592 AACGTAGCAGAAGAA---AGTAAAGCGGTGACTTTAACCGTGGATTCTGCAGACGCTGTT 2648

Qy     533 GlnAlaAspAlaMetAlaAsnAla---TyrAlaIleGluLysGlyLysArgLeuPhePhe 551
              ||||||   :::            |||               ||||||
Db    2649 GTGACCGATGCCTATTCGGGTACGGAATATACAGCAACAGCAGGAAAA----------- 2696

Qy     552 AlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMet 571
              :::                  ||||||:::|||   |||                   |||
Db    2697 ------GTAAATTTAACTCTTCCAGGAAAAGCAGATCCCGGAACGGTTTTATTGACG--- 2747

Qy     572 ArgMetSerTyrGlySerIleLysGly--------------------------------- 580
              :::      |||:::|||   |||
Db    2748 ---GTAGAAGGAGGCAACATAACCGGTGTTGCGAAAGACAACGGTGAAGTAGTGGTTGAA 2804

Qy     581 ---------------------------------TyrGluProGlnAspGlyAlaTrpTyrAsn 590
                                              |||:::   :::|||      :::   |||
Db    2805 CTAGTACCGGAAAACAATATCCGCATTCACTATAAGCGCGAAGACAATGTGTATAAAAAC 2864

Qy     591 TyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAla 610
           |||         |||      :::                     |||         :::
Db    2865 TAC---------GGCGCATGGCTATGGAATGATGTTGCGTCTCCTTCTGCAAACTGGCCA 2915

Qy     611 ValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGlu--- 629
           |||         :::   :::         :::|||   :::||||||   |||   :::
Db    2916 GTAGGGGCAACGATGTTCGAG------AAAACCGATAGTTACGGTGCGTATATCGATGTG 2969

Qy     630 ---------AsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGly 646
                     :::|||      :::|||   ||||||         |||||||||   |||
Db    2970 CCATTAGCTGACGGAGCGAAAAATATCGGCTTTCTTGTC---ATGGATATTACCGCGGGT 3026

Qy     647 AsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGly 666
           ::::::|||               ||||||         :::               :::
```

```
Db     3027 GACGCGGGCAAAGACGGCGGAGATAAAGGCTTTACGATTTCTTCACCACAAGCAAATGAA 3086

Qy      667 AsnTrpGluAlaMetSerGlyAsp------IleGluPheGluPro---AspLeuGlnArg 683
              |||               |||              :::||||||   ||||||
Db     3087 ATTTGGATCAAGCAAGGTTCGGACAAAGTGTACACGTACGAACCGGTAGATTTGCCAGCA 3146

Qy      684 ThrIleSerValAspIleArgTyrValLeuPheMetIleAsp----------------- 697
                :::|||   |||   |||             :::|||
Db     3147 AAT---ACAGTCCGCATTCATTACACGCGTGAAGCGGTAGATTACGATGACTTTGGCATA 3203

Qy      698 ---LysTrpGly 700
                |||||| 
Db     3204 TGGAATTGGGGC 3215

RESULT 9
AE008428
LOCUS       AE008428               10292 bp    DNA     linear   BCT 13-SEP-2001
DEFINITION  Streptococcus pneumoniae R6 section 44 of 184 of the complete
            genome.
ACCESSION   AE008428 AE007317
VERSION     AE008428.1  GI:15458055
KEYWORDS    .
SOURCE      Streptococcus pneumoniae R6.
  ORGANISM  Streptococcus pneumoniae R6
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 10292)
  AUTHORS   Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
            DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
            Gilmour,R., Glass,J.S., Khoja,H., Kraft,A., LaGace,R.,
            LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
            McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
            Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
            Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
            Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
            and Glass,J.I.
  TITLE     Genome of the bacterium Streptococcus pneumoniae strain R6
  JOURNAL   J. Bacteriol. 183 (19), 5709-5717 (2001)
  MEDLINE   21429245
   PUBMED   11544234
REFERENCE   2  (bases 1 to 10292)
  AUTHORS   Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
            DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
            Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., LaGace,R.,
            LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
            McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
            Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
            Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
            Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
            and Glass,J.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
            Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES             Location/Qualifiers
     source          1..10292
                     /organism="Streptococcus pneumoniae R6"
                     /strain="R6"
                     /db_xref="taxon:171101"
     repeat_region   147..341
                     /rpt_family="BOX element"
                     /rpt_type=dispersed
                     /rpt_unit=Box_ABBBC
     gene            353..598
                     /gene="spr0485"
     CDS             353..598
                     /gene="spr0485"
                     /codon_start=1
                     /transl_table=11
                     /label=spr0485
                     /product="Hypothetical protein"
                     /protein_id="AAK99289.1"
                     /db_xref="GI:15458056"
                     /translation="MEVVMDNIIDVSIPVAEVVDKHPEVLEILVELGFKPLANPLMRN
                     TVGRKVSLKQGSKLAGTPMDKIVRTLEANGYEVIGLD"
     gene            598..1932
                     /gene="spr0486"
     CDS             598..1932
                     /gene="spr0486"
                     /codon_start=1
                     /transl_table=11
                     /label=spr0486
                     /product="Conserved hypothetical protein"
                     /protein_id="AAK99290.1"
                     /db_xref="GI:15458057"
                     /translation="MTDERIHILRDILLELHNGASPESVQDRFDATFTGVSAIEISLM
                     EHELMNSDSGVTFEDVMELCDVHANLFKNAIKGVEVLDTEHPGHPVRVFKEENLALRA
                     ALIRIRRLLDTYELMEDEEMLAEMRKGLVRQMGLVGQFDIHYQRKEELFFPIMERYGH
                     DSPPKVMWGVDDQIRELFQTALTTAKSLPEVSISSVKEAFEAFATEFESMIFKEESIL
                     LMILLESFTQDDWIQIAEESDAYGYAIIRPSEKWVPERQSFIEEKIAEEPVQLDTAEG
                     QVQQVIDTPEGHFTITFTPKEKEAVLDRHSQQAFGNGYLSVEQANLILNHLPMEITFV
                     NKEDIFQYYNDNTPADEMIFKRTPSQVGRNVELCHPPKYLDKVKAIMKGLREGTKDKY
                     EMWFKSESRGKFVHITYAAVHDEDGEFQGVLEYVQDIQPYREIDTDYFRGLE"
     gene            1942..2196
                     /gene="spr0487"
     CDS             1942..2196
                     /gene="spr0487"
                     /codon_start=1
                     /transl_table=11
                     /label=spr0487
                     /product="Hypothetical protein"
                     /protein_id="AAK99291.1"
                     /db_xref="GI:15458058"
                     /translation="MSYEQEFMKEFEAWVNTQIMINDMAHKESQKVYEEDQDERAKDA
                     MIRYESRLDAYQFLLGKFENFKAGKGFHDLPEGLFGERNY"
     gene            2292..2687
                     /gene="spr0488"
     CDS             2292..2687
                     /gene="spr0488"
                     /codon_start=1
                     /transl_table=11
                     /label=spr0488
                     /product="Hypothetical protein"
                     /protein_id="AAK99292.1"
                     /db_xref="GI:15458059"
                     /translation="MSKKLNRKKQLRNGLRRSGAFSSTVTKVVDETKKVVKRAEQSAS
                     AAGKAVSKKVEQAVEATKEQAQKVANSVEDFAANLGGLPLDRAKTFYDEGIKSASDFK
                     NWTEKELLDLKGIGPATIKKLKENGIKFK"
     gene            3099..3548
                     /gene="spr0489"
     CDS             3099..3548
                     /gene="spr0489"
                     /codon_start=1
                     /transl_table=11
                     /label=spr0489
                     /product="Hypothetical protein"
                     /protein_id="AAK99293.1"
                     /db_xref="GI:15458060"
                     /translation="MIGVVARENAAEQIKQYQKFTVNISDETSMLAMEQAGFISHQEK
                     LERLGVHYEISERTQIPILDACPLVLDCRVDRIVEEDGICHIFAKILERLVAPELLDE
                     KGHFKNQLFAPTYFMGDGYQRVYRYLDKRVDMKGSFIKKARKKDGKN"
     gene            3535..4092
                     /gene="spr0490"
     CDS             3535..4092
                     /gene="spr0490"
                     /note="Probable acetyl transferase"
                     /codon_start=1
                     /transl_table=11
                     /label=spr0490
                     /product="Conserved hypothetical protein"
                     /protein_id="AAK99294.1"
                     /db_xref="GI:15458061"
                     /translation="MARTELPDKIETERLVLRVRTVADAEDIFDYASLPEVAYPAGFP
                     PVKTLEDEIYYLEHILPERNQKENLPAGYGIVVKGTDKIVGSVDFNHRHEDDVLEIGY
                     TLHPDYWGRGYVPEAARALIDLAFKDLGLHKIELTCFGYNLQSKRVAEKLGFTLEARI
                     RDRKDVQGNRCDSLIYGLLKSEWEE"
     gene            4094..5056
```

```
                     /gene="spr0491"
     CDS             4094. .5056
                     /gene="spr0491"
                     /codon_start=1
                     /transl_table=11
                     /label=spr0491
                     /product="Hypothetical protein"
                     /protein_id="AAK99295.1"
                     /db_xref="GI:15458062"
                     /translation="MSDVKEEVSSLSEKQLRQIDVEYAELNDSDIIERLAYLEINNNE
                     KRIVISDIEPTKEIMSVSDQIFEIQKNFQKIKNMFELFISDVSDFLSIKNKLESKELE
                     IEEADVNRFMIHLLSSGKLFVDFNENQIKQKYSKDSEEFDCIHGFASYQYDINFTYRF
                     CHSLRNYSQHTDLPINEVKAVSPDDETVIIDFYIDLDYLLNSNFKWKKLKGELIKLNQ
                     ETSKIDAIALVKEYFNALTELYGNYNKLFLFKLNHNTLVDIKSKLESLKLKHSRYYISK
                     ISKYDLKYNPGNYTMSPLAAFAEIEEIYIELSKIGLVKIVNKSN"
     gene            5068. .7719
                     /gene="valS"
                     /note="spr0492"
     CDS             5068. .7719
                     /gene="valS"
                     /EC_number="6.1.1.9"
                     /codon_start=1
                     /transl_table=11
                     /label=spr0492
                     /product="Valyl-tRNA synthetase"
                     /protein_id="AAK99296.1"
                     /db_xref="GI:15458063"
                     /translation="MSKELSSKYNPAEVEAGRYQKWLDADVFKPSGDQKAKPYSIVIP
                     PPNVTGKLHLGHAWDTTLQDIIIRQKRMQGFDTLWLPGMDHAGIATQAKVEERLRGEG
                     ISRYDLGRESFLTKVWEWKDEYATTIKEQWGKMGLSVDYSRERFTLDEGLSKAVRKVF
                     VDLYKKGWIYRGEFIINWDPAARTALSDIEVIHKDVEGAFYHMNYMLEDGSRALEVAT
                     TRPETMFGDVAIAVNPEDPRYKDLIGKNVILPIANKLIPIVGDEHADPEFGTGVVKIT
                     PAHDPNDFLVGQRHNLPQVNIMNDDGTMNELVFEFSGMDRFEARKAVVAKLEEIGALV
                     KIEKRVHSVGHSERTGVVVEPRLSTQWFVKMDQLAKNAIANQDTEDKVEFYPPRFNDT
                     FLQWMENVHDWVISRQLWWGHQIPAWYNADGEMYVGEEAPEGDGWTQDEDVLDTWFSS
                     ALWPFSTMGWPEVDSEDFKRYFPTSTLVTGYDIIFFWVSRMIFQSLEFTGRQPFQNVL
                     IHGLIRDEQGRKMSKSLGNGIDPMDVIEKYGADALRWFLSNGSAPGQDVRFSYEKMDA
                     SWNFINKIWNISRYILMNNGGLTLDVAHDNVTKVATGEAGNVTDRWILHNLNETIAKV
                     TENFDKFEFGVAGHILYNFIWEEFANWYVELTKEVLYSDNEDDKVITRSVLLYTLDKI
                     LRLLHSIMPFVTEEIFGQYAEGSIVTAAYPTVNPAFEDLAAHTGVESLKDLIRAVRNA
                     RAEVNVAPSKPITILVKTSDSDLEAFFNSNVNYIKRFTNPEHLEIASTIPAPELAMSS
                     VITGAEIFLPLVDLLNVEEELARLEKELAKWQKELDMVGKKLSNERFVANAKPEVVQK
                     EKDKQADYQAKYDVTVARIDEMKKLVK"
     gene            7774. .8544
                     /gene="spr0493"
     CDS             7774. .8544
                     /gene="spr0493"
                     /codon_start=1
                     /transl_table=11
                     /label=spr0493
                     /product="Conserved hypothetical protein"
                     /protein_id="AAK99297.1"
                     /db_xref="GI:15458064"
                     /translation="MMEKVMSKHPHYELLNLIGYGLAKFDKLFIKEFQCSSKSEFYRY
                     VVSLGIAETTGVVKNRMDLFDPYFDNNRKGWWQKAEVYRFRKDLIDMMFGNEDVHSYA
                     EIVKMLLASEGKKTGITIVEKPIVRTKFKRLQETGMEAENYFILHFDKEEKFQGGQLT
                     DARLYGDGYDFQVDVQEYSYLAEVKGIRKSKGRVRLTAKEFEKVKEFQSDFILSLVTN
                     LDDIPKLVLIDNPLKHFEFKKNIIKNEIIEYRSVEDLY"
     gene            8827. .9630
                     /gene="spr0494"
     CDS             8827. .9630
                     /gene="spr0494"
                     /codon_start=1
                     /transl_table=11
                     /label=spr0494
                     /product="Hypothetical protein"
                     /protein_id="AAK99298.1"
                     /db_xref="GI:15458065"
                     /translation="MQPTYNIDNPNLSYEAKRDLWRIGFGLQKVDNLVPSAYMESLAE
                     KQSRGELTYEQVYEDATAYHHTIDASTEEADLVSLRIVELLSRRGFSFSPATLLAIHK
Alignment Scores:
Pred. No.:              18.3          Length:          10292
```

```
Score:                    133.00          Matches:         106
Percent Similarity:       33.20%          Conservative:    64
Best Local Similarity:    20.70%          Mismatches:      164
Query Match:              3.58%           Indels:          178
DB:                       1               Gaps:            25

US-10-008-355-2 (1-712) x AE008428 (1-10292)

Qy     192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
           |||||||||                       |||         :::|||    |||
Db    5143 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 5202

Qy     204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
           |||:::    ||||||                                       |||
Db    5203 CCAAACGTTACAGGTAAACTT---------------------------------CACCTT 5229

Qy     224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
           |||                                                   ::::::
Db    5230 GGT---------------------------------------------------CACGCT 5238

Qy     244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
              |||         :::   :::                  ||||||||| :::
Db    5239 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAAACGCATGCAAGGTTTTGAT 5295

Qy     264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                             |||:::    ||||||                  :::   |||
Db    5296 -----------------ACCCTTTGGCTTCCTGGGATGGACCACGCAGGGATTGCCACT 5337

Qy     283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                    ||||||:::|||:::     |||       |||   ::::::
Db    5338 CAGGCTAAGGTTGAGGAGCGCTTGCGTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT 5397

Qy     303 ------------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                       |||      ||||||:::                  :::   ||||||
Db    5398 GAGTCTTTCTTGACGAAAGTCTGGGAATGGAAAGAC------------GAATATGCCACT 5445

Qy     319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
              ||||||         |||        |||   :::|||   :::
Db    5446 ACCATCAAGGAACAATGGGGCAAGATGGGGCTCTCTGTAGACTATTCTCGTGAGCGTTTC 5505

Qy     338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
              ::::::   ||||||:::                       :::|||      :::
Db    5506 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 5541

Qy     358 AlaPheAlaAspTrpIleArgLys------------AsnGlyLysSerAlaValTyrGly 373
              |||   |||      :::|||                |||:::      :::
Db    5542 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 5601

Qy     374 Asp---------------ValLeuSerSerLeuGluLysAlaTyrLys------------ 384
           |||               ||||||   :::|||      :::|||
Db    5602 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTGGAAGGT 5661

Qy     385 ------------------------------GluGlyAlaLysAlaAsnArgGluMetThr 394
                                         :::|||:::::::|||           |||
Db    5662 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGCCCTTGAAGTTGCTACA 5721

Qy     395 TyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                    ||||||:::||||||
Db    5722 ACTCGTCCTGAGACTATGTTTGGGGACGTT----------------------------- 5751

Qy     415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
           |||:::|||   ||||||:::                           |||   :::|||
Db    5752 GCGATTGCGGTTAATCCAGAA---------------------------GACCCGCGCTAC 5784

Qy     435 LysAsp------------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
           ||||||                  ||||||   :::         |||:::::::|||
Db    5785 AAGGACTTGATTGGTAAAAACGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 5841

Qy     449 MetLeuAsp------------------------------------IleValArgArg 455
           :::   |||                                    |||
Db    5842 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 5901

Qy     456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
```

```
                |||    |||    |||                         |||:::
Db     5902 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACATCATG 5961

Qy      476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
              |||     .        :::   ||||||:::   |||   :::          |||
Db     5962 AACGACGACGGAACTATGAACGAGCTTGTCTTTGAATTTTCAGGCATG---------GAT 6012

Qy      496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
            :::|||   |||                     |||     ||||||      |||
Db     6013 CGTTTTGAAGCTCGT---------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 6057

Qy      516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla------------ 531
                     |||:::::   |||   |||   :::         :::
Db     6058 GGTGCCCTTGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAACGTACAGGT 6117

Qy      532 ---------------------------------------------------IleGlnAla 534
                                                                :::
Db     6118 GTCGTAGTTGAGCCACGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 6177

Qy      535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
            :::|||:::||||||             :::|||
Db     6178 AACGCCATTGCCAACCAAGACACAGAGGACAAG--------------------------- 6210

Qy      555 ArgGluMetTyrProGlyArgAlaLeuProSer--------------------------- 565
               |||   ||||||   |||             :::
Db     6211 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 6270

Qy      566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
            |||         :::   ::::::   :::|||                         |||
Db     6271 GACTGGGTTATCTCTCGTCAGCTCTGGTGGGGTCAC------------------CAAATC 6312

Qy      586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
               |||||||||||            |||:::      :::   ::::::   |||:::
Db     6313 CCTGCTTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 6366

Qy      606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
               |||   :::         |||:::::::|||||
Db     6367 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 6402
```

RESULT 10
SPNEU1923/c

```
LOCUS       SPNEU1923        11362 bp    DNA     linear   HTG 11-JUL-2001
DEFINITION  Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
            in ordered pieces.
ACCESSION   AL449945
VERSION     AL449945.1  GI:11545170
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 11362)
  AUTHORS   Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
            Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
            Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
            and Garcia-Bustos,J.F.
  TITLE     Annotated draft genomic sequence from a Streptococcus pneumoniae
            type 19F clinical isolate
  JOURNAL   Microb. Drug Resist. 7 (2), 99-125 (2001)
  MEDLINE   21335329
   PUBMED   11442348
REFERENCE   2  (bases 1 to 11362)
  AUTHORS   Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
            Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
            Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
            Garcia-Bustos,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
            Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT     * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..11362
                     /organism="Streptococcus pneumoniae"
                     /serotype="19F"
                     /db_xref="taxon:1313"
                     /clone="G54"
BASE COUNT     3304 a   2518 c   2123 g   3414 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:              20.6        Length:         11362
Score:                  133.00      Matches:        116
Percent Similarity:     33.62%      Conservative:   77
Best Local Similarity:  20.21%      Mismatches:     194
Query Match:            3.58%       Indels:         187
DB:                     2           Gaps:           26

US-10-008-355-2 (1-712) x SPNEU1923 (1-11362)

Qy      131 LysIleValLysValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMet 150
            ||||||                                 |||      |||  |||  ::::::::
Db     5039 AAAATAAATAATACGAATTTGTCTGCAGAAGAAGCTGCTAAGCAAATTCAAAATAAAATA 4980

Qy      151 GluArgLeu-ArgLysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAs 170
               :::::: ||||||          :::   :::||||||:::   |||   |||
Db     4979 AACAAAATAGAGAAAGGACACACACATGTCTAAAGAACTTTCATCTAAATACAATCCAGC 4920

Qy      170 pGluAsnGlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyrPheLeuIleVa 190
             :::   :::          :::   :::                        :::
Db     4919 CGATGTTGAGGCTGGTCGTTACCAAAAATGG-----------------------CTTGA 4884

Qy      190 lTyrAspValPhe-----------------------LysAspValArgMetValPheAl 202
                |||||||||                        |||       :::|||
Db     4883 TGCGGATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTGATTCC 4824

Qy      202 aProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHi 222
              ||||||:::      ||||||                                  ||
Db     4823 ACCACCAAACGTAACTGGGAAACTT---------------------------------CA 4797

Qy      222 sThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTy 242
             |   |||                                                   ::
Db     4796 CCTTGGT---------------------------------------------------CA 4788

Qy      242 rSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTy 262
             :::   |||          |||   :::   :::            |||||||||::
Db     4787 CGCTTGGGATACG---ACTTTGCAGGATATCATCATCCGTCAAAAACGCATGCAGGGCTT 4731

Qy      262 rLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu-- 281
             :                 |||:::      ||||||            :::
Db     4730 TGAT-----------------ACCCTTTGGCTTCCAGGTATGGACCACGCTGGTATTGC 4689

Qy      282 -ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValAr 301
              |||           ||||||:::|||:::      |||         |||   :::::
Db     4688 GACTCAGGCTAAGGTTGAGGAGCGCTTACGTGGTGAGGGTATTACGCGTTATGACCTAGG 4629

Qy      301 gGly-----------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAl 317
                              |||        ||||||:::                 :::   ||
Db     4628 TCGTGAGTCTTTCTTGACGAAGGTCTGGGAATGGAAAGAC------------GAATATGC 4581

Qy      317 aThrArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSe 336
             ||||   ||||||           |||         :::|||     :::
Db     4580 CACTACCATCAAGGAACAATGGGGCAAGATGGGGCTTTCTGTAGACTATTCTCGTGAGCG 4521

Qy      336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGl 356
                   ::::::   ||||||:::                          :::|||.
Db     4520 TTTCACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCG 4485

Qy      356 uArgAlaPheAlaAspTrpIleArgLys-----------AsnGlyLysSerAlaValTy 372
             :::   |||    |||      :::|||             |||:::       :::
Db     4484 TAAGGTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTATCGTGGTGAGTTTATCATCAA 4425

Qy      372 rGlyAsp--------------ValLeuSerSerLeuGluLysAlaTyrLys-------- 384
                |||           .   ||||||   :::|||      :::|||
```

```
Db     4424 CTGGGACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTAGA 4365

Qy      385 ---------------------------------GluGlyAlaLysAlaAsnArgGluMe 393
                                             :::|||:::::|||
Db     4364 AGGTGCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGCCCTTGAAGTTGC 4305

Qy      393 tThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAl 413
              |||          ||||||:::||||||
Db     4304 TACAACTCGTCCTGAGACTATGTTTGGGGACGTT------------------------- 4271

Qy      413 aAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLy 433
                 |||:::|||  ||||||:::                          |||  ::
Db     4270 ----GCGGTTGCGATTAATCCAGAA--------------------------GACCCGCG 4242

Qy      433 sTyrLysAsp-----------------TyrLeuProSerLeuAspArgLysValLeuPr 447
             :||||||||                  ||||||   :::      |||:::::||
Db     4241 CTACAAGGACTTGATTGGTAAAAACGTCATCCTTCCA---ATCGCTAATAAACTCATCCC 4185

Qy      447 oAlaMetLeuAsp---------------------------------------IleValAr 454
             |   :::  |||                                       |||
Db     4184 AATCGTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACC 4125

Qy      454 gArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPh 474
                     |||  |||   |||               |||:::
Db     4124 TGCCCACGRTCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGT 4065

Qy      474 eLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSe 494
                |||               :::    |||:::   |||    :::
Db     4064 CATGAACGACGACGGAACTATGAATGAGCTTGCCTTTGAATTTTCAGGCATG-------- 4013

Qy      494 rAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGl 514
              |||:::|||   |||               |||           ||||||     ||
Db     4012 -GATCGTTTTGAAGCTCGT---------------AAGGCAGTCGTTGCTAAGTTGGAAGA 3969

Qy      514 uLysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla-------- 531
             |            |||::::::  |||   |||   :::        :::
Db     3968 AATCGGTGCCCTTGTCAAAATCGAAAAACGTGTCCATTCAGTCGGTCACTCAGAACGGAC 3909

Qy      532 ---------------------------------------------------IleGl 533
                                                                :::
Db     3908 AGGTGTCGTAGTTGAGCCACGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGC 3849

Qy      533 nAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGl 553
                :::|||:::||||||          :::|||
Db     3848 TAAGAACGCCATTGCCAACCAAGACACAGAGGACAAG---------------------- 3812

Qy      553 yLeuArgGluMetTyrProGlyArgAlaLeuProSer---------------------- 565
                 |||   ||||||   |||          :::
Db     3811 ----GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGT 3756

Qy      566 ----AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGl 584
                 |||            :::   ::::::   :::|||                 ||
Db     3755 CCACGACTGGGTTATCTCTCGTCAGCTTTGGTGGGGTCAC-----------------CA 3714

Qy      584 nAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspPr 604
             |     ||||||||||||              |||:::      :::    ::::::  ||
Db     3713 AATCCCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCC 3660

Qy      604 oLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
             |:::    |||     :::          |||::::::||||||
Db     3659 AGAAGGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 3620
```

```
RESULT 11
AE009858
LOCUS       AE009858             14169 bp    DNA     linear   BCT 16-JAN-2002
DEFINITION  Pyrobaculum aerophilum strain IM2 section 113 of 201 of the
            complete genome.
ACCESSION   AE009858 AE009441
VERSION     AE009858.1  GI:18160615
KEYWORDS    .
SOURCE      Pyrobaculum aerophilum.
  ORGANISM  Pyrobaculum aerophilum
            Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
            Thermoproteaceae; Pyrobaculum.
REFERENCE   1  (bases 1 to 14169)
  AUTHORS   Fitz-Gibbon,S.T., Ladner,H., Kim,U.J., Stetter,K.O., Simon,M.I. and
            Miller,J.H.
  TITLE     Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
            aerophilum
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (2), 984-989 (2002)
   PUBMED   11792869
REFERENCE   2  (bases 1 to 14169)
  AUTHORS   Fitz-Gibbon,S.T., Ladner,H., Kim,U.-J., Stetter,K.O., Simon,M.I.
            and Miller,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2001) Microbiology and Molecular Genetics,
            University of California, Los Angeles, 405 Hilgard Ave, Los
            Angeles, CA 90095-1489, USA
FEATURES             Location/Qualifiers
     source          1..14169
                     /organism="Pyrobaculum aerophilum"
                     /strain="IM2"
                     /db_xref="taxon:13773"
     gene            144..1019
                     /gene="PAE2116"
     CDS             144..1019
                     /gene="PAE2116"
                     /note="Protein fate; Protein modification and repair"
                     /codon_start=1
                     /transl_table=11
                     /product="methionine aminopeptidase"
                     /protein_id="AAL63956.1"
                     /db_xref="GI:18160616"
                     /translation="MLRILRQVGDVVNKALKYALDLTQPDMPVLELCERVEAFIRAND
                     AKPAFPVNVSVNEVAAHYTAKRGDSLKIPKTGLVKIDVGAQRDGYIVDAAVTVTLGPV
                     FNNLQKAAKSALESAINAAKPGIKAWQIGEVVERVIKSFGLSPIYNLTGHKVERYLLH
                     AGHVIPNYSDKTASQALAPGDVYAIEPFATNGEGYVVDGGEITIYRLARMRHKNYQHL
                     IDIISAEAGPLPFTPRWFPQLTDDAIREALKAGVLQGYEVLVERSKGFVAQFEDTVYV
                     GENEVIPLVRTLELL"
     gene            1060..2709
                     /gene="PAE2117"
     CDS             1060..2709
                     /gene="PAE2117"
                     /note="Protein fate; Protein folding and stabilization"
                     /codon_start=1
                     /transl_table=11
                     /product="thermosome (chaperonin) alpha subunit"
                     /protein_id="AAL63957.1"
                     /db_xref="GI:18160617"
                     /translation="MSQAVLTQIGGVPVLVLKEGTQRAFGKEALRLNIMIARAIAEVM
                     RTTLGPKGMDKMLIDSLGDITITNDGATILDEMDVQHPIAKLLVEISKSQEEEAGDGT
                     TTAVVLAGALLEEAEKLLEKNIHPTVIVSGFKKALDVAAEHLRKVAIPVNRTDVDTLR
                     KIAMTSMGGKISETVKEYFADLAVKAVLQVAEERNGKWYVDLDNIQIVKKHGGSLLDT
                     QLVYGIVVDKEVVHAAMPKRVVNAKIALLDAPLEVEKPEIDAEIRINDPTQMRAFLEE
                     EERILRGYVDKLKSLGVTVLFTTKGIDDIAQYYLAKAGILAVRRVKRSDIEKLVRATG
                     ARLVTSIEDLTEADLGFAGLVEERRVGDEKMVFVEQCKNPKAVSILVRGGFERLVDEA
                     ERNLDDALSVVSDVVEDPYILPAGGAAEIEAAKAVRAFAPKVGGREQYAVEAFARALE
                     VIPKALAENAGLDPIDILTELTHKHEQTDGWKYGLDVYQGKVVDMVSLGLVEPLTVKI
                     NALKVAVEAASMILRIDEIIAASKLEKEKEEKKEEKKEEFD"
     gene            complement(2706..3257)
                     /gene="PAE2118"
     CDS             complement(2706..3257)
                     /gene="PAE2118"
                     /note="Hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAL63958.1"
                     /db_xref="GI:18160618"
                     /translation="MLAEYVAQRVALIQERAYHRLVVILSATSMSSVMMSGLALFFEP
                     AFKYLSLALLALGVLLSRRVLRNAEEKIRKFMANVEREAMARRYFATLKEYYNPIDYL
                     TKLYAEPTIEDPIIAAEKKKIRASTYLFSQSITALRAKNRLVITITALTISIPHLVLN
                     KDIIAIGVILLIVILLFRKRSWV"
     gene            complement(3247..4278)
                     /gene="PAE2119"
     CDS             complement(3247..4278)
```

```
                     /gene="PAE2119"
                     /note="Unclassified"
                     /codon_start=1
                     /transl_table=11
                     /product="conjugal transfer protein, conjectural"
                     /protein_id="AAL63959.1"
                     /db_xref="GI:18160619"
                     /translation="MYPILKRLLECAECKSSCVEKGICDFAKEELPLVIKLLSRVFKK
                     SIDETLDFRYGVLRRINEKMAIRGVLATVGLEELSPFLEDREVEDIVVIPGRPIYITR
                     KSGKEKADVTAELKTVKSLLKIAYLKGVELTTSNPSLRYGLKLGDLRLRISLDLPPVV
                     PVPQAYIRIHRGKIAVADMLKSGFITEEQLKTVYAWVKEGRHIVITGPPGSGKTTLLS
                     VIDDLIPGQLQRVYIDEADEFDDDPDKNQIKIRNVNKLREVYASLNRNIDIIIGELQ
                     YEEHFNAFKTAVEIGLQTLATMHATSVKDALKRLERRGISRENLAIIQLAKRYDEGIE
                     RRVVELYAS"
     gene            complement(4250..4729)
                     /gene="PAE2120"
     CDS             complement(4250..4729)
                     /gene="PAE2120"
                     /note="Transcription; Transcription factors"
                     /codon_start=1
                     /transl_table=11
                     /product="transcription initiation factor IIB,
                     conjectural"
                     /protein_id="AAL63960.1"
                     /db_xref="GI:18160620"
                     /translation="MKCPYCQSMNITLADGEYVCRDCGTVLGPEVVPPRPRQVLPVPV
                     RHKLIMLALEQENKKSIKKKYSEMVKMYTGKVAEALGVPEVTVVALQIFQRLDKRIYQ
                     GKSPRVIAAAVAYLAAERLGIYIHKQVIAKILGVSKFSIRDTASRLRKYVPNIKETT"
     gene            4857..5171
                     /gene="PAE2121"
     CDS             4857..5171
                     /gene="PAE2121"
                     /note="Hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAL63961.1"
                     /db_xref="GI:18160621"
                     /translation="MEFSLAVVDVTNDMPNVIKLKGDVEATIELPLEKLGVVFSVGDR
                     LKLVLQKEKDQDLNKYKVYAWGIVYYVGNGMTRISIGGLQLDIMKELPLKAGEKIYVG
                     IL"
     gene            5207..5641
                     /gene="PAE2122"
     CDS             5207..5641
                     /gene="PAE2122"
                     /note="Transcription; RNA processing"
                     /codon_start=1
                     /transl_table=11
                     /product="small nuclear ribonucleoprotein homolog
                     (Sm-like)"
                     /protein_id="AAL63962.1"
                     /db_xref="GI:18160622"
                     /translation="MASTLAEASKRFVAELNNLLGREVQVVLSNGEVYKGVLHAVDNQ
                     LNIVLANASNKAGEKFNRVFIMYRYIVHIDSTERRIDMREFAKQAEKIFPGMVKYIEE
                     TNVVLIGDKVRVSEIGVEGVGPVAERAKRLFEEFLKRYSLEQ"
     gene            complement(5871..6710)
                     /gene="PAE2125"
     CDS             complement(5871..6710)
                     /gene="PAE2125"
                     /note="Hypothetical; Conserved within genome"
                     /codon_start=1
                     /transl_table=11
                     /product="paREP2a"
                     /protein_id="AAL63963.1"
                     /db_xref="GI:18160623"
                     /translation="MTVKAVKAACVKVALPEAFSLPRKFGVSWGCFTQGAFYGVDAVI
                     AEGGRKGIRFRYVLPYDTIAKRYMAEYKPKKRWEGYGISPTLAEALYELAENHKIEIM
                     IEKTRYKTYYYAVVDGVRIEGSLCYRESLRGCISEVLRRVKEQREWEEQLRRQRIREE
                     VLNALTEWYFRCFGETPAHPLKVAKRLEEYHALCQMRSAAKREFGIAPMERALRGAFW
                     WEGEWRGKSMSCFVTEKRAVCKVGEKMAAFYVFDTPHGVYLRPEIKLVDDWIKVAHRG
                     DDS"
     gene            complement(6700..7077)
                     /gene="PAE2127"
     CDS             complement(6700..7077)
                     /gene="PAE2127"
                     /note="Hypothetical; Conserved within genome"
                     /codon_start=1
                     /transl_table=11
                     /product="paREP2a"
                     /protein_id="AAL63964.1"
                     /db_xref="GI:18160624"
                     /translation="MAALKDWYRRCFKWPIMPGEEGKLVRRIELYYGMCEMAKTAIAE
                     YGEKYAEPLISEYALRKAFWWEGEWRGKPMSCFVTEKKAVCKVGDKMATFYVFDTPHG
                     VYLRPEIKLIDDWIKVAYRGDDS"
     gene            7235..8263
                     /gene="PAE2128"
     CDS             7235..8263
                     /gene="PAE2128"
                     /note="Hypothetical; Conserved within genome"
                     /codon_start=1
                     /transl_table=11
                     /product="paREP2b"
                     /protein_id="AAL63965.1"
                     /db_xref="GI:18160625"
                     /translation="MEEAPGCIAAVEEAVKPAVVAAAGLAGYAALHDGLYSTAVVSAT
                     AAAVVLAREGAFERAVEYVRRAAEAAYEAAREVFEKARITLQRLYELFVEAVARVLDY
                     VRAHWFIMAAAAAGLMAWVVAQQLDYTLWQNHIALNAGAIAGLAKAAGVGDKWKEVGN
                     AVLTKTASEKEIVERAAQLGLDKGVVEKAVSAFSTLRHAMDEKQIENAVLELRRLTSY
                     VQAMKAANYKKALEEASKAWYEQRDPAALWALAVLGARETGKSVKVFEDDDDRIAYLI
                     ASFTLLNHLREFVELKNKSYKALERLEGAVKSGRFALSDVWGDLNVLVEAAERAKGGY
                     QRAGQNCI"
     gene            8307..11282
                     /gene="PAE2129"
     CDS             8307..11282
                     /gene="PAE2129"
                     /note="Hypothetical; Conserved within genome"
                     /codon_start=1
                     /transl_table=11
                     /product="paREP2b"
                     /protein_id="AAL63966.1"
                     /db_xref="GI:18160626"
                     /translation="MATENVIRLAEATRDDLPKIDATRGERTVTALLSLVTGGLFGVT
                     VERALAERGKWTVGALGIASAVRSTPSNFYNAFSSVSGDEKPSEAKKLTFRIATLLAD
                     PAMREILAGRQGVEVRVEQKTENGKKYIYITFVDKDGHELLKAPWDADRGLMPLWAEG
                     EVVELFKEVGKLAVAVSSGSKPLEELNEEEWKKVVETVRRVKKAIEWTVKAVAIGALP
                     TDAVLLPGSEYVRGTSSYLSQTFTYWALAKDRIGLVRVYPSEEGLKPMWRVEGKYTEA
                     VREVLSVSRAVFENLLTPNKQSEGGVDLKVALTDVKMNDELKKALETAAGEFWERVNE
                     LLALWKQLEEEAKEKKDEERKKVVMEVDKLGKYLRVLLPLAYAVEARKMGELSREEAA
```

```
Alignment Scores:
Pred. No.:                 34.2           Length:          14169
Score:                     131.50         Matches:         138
Percent Similarity:        34.63%         Conservative:    112
Best Local Similarity:     19.11%         Mismatches:      247
Query Match:               3.54%          Indels:          225
DB:                        1              Gaps:            36

US-10-008-355-2 (1-712) x AE009858 (1-14169)

Qy      111 ValSerArgThrMetGlyGluGluLeuProIlePro-----------------GlyLeu 124
            ||||||||   :::   |||  |||   |||                      |||:::
Db     9153 GTAAGCCGCGCCGTGTTTGAAAATCTATTGACACCTAACAAACAGTCGGAAGGCGGCGTT 9212

Qy      125 SerValLysTyrLeuArgLysIleValLysValThrAspLysValGluGlyGlnLeuLys 144
               :::|||             ||||||:::  |||:::::::::      |||:::
Db     9213 GATCTGAAGGTGGCTCTAACTGATGTTAAAATGAATGATGAATTAAAGAAGGCCTTGGAG 9272

Qy      145 GlyIleThrAspGluMet---GluArgLeuArgLysAlaGlnGluValCysGlnGluLeu 163
                     |||      ||||||:::   :::          :::   ::::::|||
Db     9273 ACGGCGGCCGGCGAGTTCTGGGAGCGCGTTAATGAGCTGTTAGCACTGTGGAAACAGCTA 9332

Qy      164 AlaLys------LysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyr 181
               :::      ||||||      ||||||   :::         :::|||
Db     9333 GAAGAGGAGGCAAAGGAGAAAAAGGACGAAGAGAGGAAAAAGGTGGTAATGGAGGTAGAC 9392

Qy      182 SerAsnAsnGluTyr------------------------------------------- 186
```

```
                           :::|||
Db     9393 AAGCTGGGCAAATACCTCCGCGTCCTCCTCCCGCTGGCATATGCCGTAGAGGCGCGCAAG 9452

Qy      187 ------------------------------------------PheLeuIleValTyrAsp 192
                                                      |||   ::::::||||||
Db     9453 ATGGGTGAACTGTCTAGAGAGGAGGCGGCCCTTGCGGTGATTTTCGCTGTGTTGTACGAT 9512

Qy      193 ------ValPheLysAspValArgMetValPhe-----------------------Ala 202
                  ||||||:::|||    ::::::|||                          :::
Db     9513 GGGACTGTGTTTAGAGACGAAATTCTGCTTTTCATCGGCGGCCCTGAACACGTGGAAAGC 9572

Qy      203 ProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHis 222
            |||      :::        |||       |||      |||:::|||
Db     9573 CCGATTATGACCCGCGACCACTTC---------ACCGCTTTCTGGCTGTGGGCTCTGAAG 9623

Qy      223 ThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyr 242
                                       ||||||   |||         |||   |||   :::
Db     9624 GAGCTTGGCCTTAAGCCAAGCTCAGTGTATAGGGGC---------AGAGGCGCGCACCAT 9674

Qy      243 SerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyr 262
                                              :::   |||         :::   |||
Db     9675 --------------------------ATTGCCTTCAAAGGCGCCGAGCTGAACGGCTTA 9707

Qy      263 LysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAsp------------ 278
                           |||:::|||          |||         :::
Db     9708 CTGGAG---------GCTGTAACGCCGGCGCTACCCGCGTTGCACGAGCTTAGGGATGCC 9758

Qy      279 ------------------------------------------------------ArgTyr 280
                                                                  ::::::
Db     9759 CTGACAGAATTCGCCGACGCGTTTAAAGTCGTTACCCGCGAGGCGGTTAAAAGGAAATTC 9818

Qy      281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300
                     ||||||   :::   |||   |||                     ||||||
Db     9819 GGCATTGACTGGGGTTACGACGTGAGAAACGAAATGTTTTTCAAAAAGCTAGAGGAAGTT 9878

Qy      301 ArgGlyIleLysGlnGly---IleTrpLysGlu---AlaMetSerAlaAspGln------ 316
                :::   :::      :::::::::         :::      ||||||
Db     9879 GTTACAATGGCAGAGGACTACGTCTACAGAAACGTCACAGTGGAGAGGGATCAGCTGGAC 9938

Qy      317 -----------------AlaThrArgIleLysTyr---------AlaSerLysTyrAla 327
                             |||   |||::::::             :::   |||
Db     9939 ACAAGCGGACAGTTGCCAAAGACTGTCATCCGCTTTAAGCTTGGCGGGGAGGAGGTGGCG 9998

Qy      328 GlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAsp 347
                        ||||||   |||             :::   ||||||:::   |||
Db     9999 TACATAACCGTATACTGGACTGGCAGT-----------AAGCTACTAGCCCAATTCGAC 10046

Qy      348 ValIleGlyArgLysArgAlaGluGluArgAlaPhe----------------------- 359
                   |||::::::||||||:::   |||
Db    10047 GGC---TCTCGAGAAAAAGCCGAACAGTTAGCCTCTATCATTACGGCTTTGGGCGGCAAG 10103

Qy      360 AlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeu 379
            |||:::         ||||||   |||      |||:::
Db    10104 GCGGAGGTCAAACCAAAGAATGCCAAATGGACTGTTCAC-------------------- 10142

Qy      380 GluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSer------ 397
                :::   :::|||         |||   |||      ::::::|||:::
Db    10143 ------CTCCACACCGACGGCATAATTGCCATTCGCCATGACAGCTGGTTAAATGCTGTG 10196

Qy      398 -----------GluThrLeuPhe------GlyGlyThrGluValValArgPheAlaGln 411
                       :::   |||:::         |||            |||:::   |||
Db    10197 AAGGCTTTGTTGACGAACTTTACAGTAAGGGGCTGATCAGCGAGGATAGATACGAACAG 10256

Qy      412 PheAlaAsnAlaLeuAlaThrAsnProAspAla------HisAlaGlyIleLeuLysSer 429
                     :::||||||   |||:::          :::||||||:::         |||
Db    10257 TTGGTGAGGGACATAGCTACAGGCCCCAACACCGTCAACTATGCCGGGGTGGAGTTCTCG 10316

Qy      430 LeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMet 449
            :::      |||   :::                |||:::   :::      |||
Db    10317 GTTTATCACAAGACGGAAGGTAAATACGATTTTTTAGAAATAAGGTATCACCCAGGC--- 10373

Qy      450 LeuAspIleValArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsnVal 469
                                                     :::      ||||||

Db    10374 -------------------------------------AACGAGGCCTCTAAAAATGCC 10394

Qy      470 IleAspLysLysPheLysGly--------------------------AspThrLysLys 480
            :::         |||                                      |||:::
Db    10395 GTTGTGAACGCCCTAAAGGCAAGAGGGCTGAAAGAGGGTGTACACTTCACTGTGAAAGAG 10454

Qy      481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500
            |||   |||:::            :::|||      |||:::               |||:::
Db    10455 TACGGCGACTACGAAATTCACGTGGCTGTTGAATCGTATACAAAAGCTTTGGAAGCCTTG 10514

Qy      501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520
               :::|||   ::::::         ||||||:::
Db    10515 GCACGTAGCTGGCTGAGAAAGGGAGAGTACTACGCCATTGAT------------------ 10556

Qy      521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540
               :::      :::   :::   :::|||         |||||||:::   |||
Db    10557 ---GACAGAAAGCGTGTAATCAGCGTCAAGGCGGAGCACAAAGACGCAGTTGTAAATACC 10613

Qy      541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560
                                   |||      |||||||||   |||
Db    10614 ----------------------CTAAAAACCGCTGGGCTGGAGGAG------------ 10637

Qy      561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580
                           |||   :::|||||||:::::
Db    10638 ---------------GATAGACACTTTACGGTGAAGTGG--------------------- 10661

Qy      581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600
                           ||||||   :::|||            |||   :::|||:::   |||
Db    10662 ------------GACGGA---TACTACGTAATTCGTATTACATATGAAGGCCTCCGCGAA 10706

Qy      601 LysGln-----------------AspProLysSerAspGluPheAlaValGlnGluAsn 614
               |||                 |||   :::::::::   |||
Db    10707 ATCCAGCGTATGGCGCTAAACGGCGACGTGGAGGCTGAGCGCTTTATA------AGGGAG 10760

Qy      615 IleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHis 634
            :::   |||:::   |||   :::   :::|||:::   |||            :::
Db    10761 CTTGAGGATGTGCTGAGG---CGCCGCCACGGCCAAAACGCAGTGAAAAGGCTGATTGAA 10817

Qy      635 IleAlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerPro------Val 652
            :::         :::      :::         |||         |||      |||
Db    10818 GTGTTGACGCCGGCGAGGGAGGAG---------GGGACGATAGATTTACCGCTAGAAGTG 10868

Qy      653 PheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSer 672
               |||:::   |||   ::::::               |||   :::|||   :::
Db    10869 CGCGACGAAAGGGGCAACGTGGTAGCCCGCGTTGTGGACTTGAGGTATGAATTCGTAATG 10928

Qy      673 GlyAsp---------------------------------------------IleGlu 676
            ||||||                                             :::|||
Db    10929 GGCGATCAACCCGTTGGCCAGTGCGCCGGGGAGGACTGCCGCCTGCGCATAATAGTGGAG 10988

Qy      677 PheGluProAspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIle 696
            :::|||   :::   :::|||                  ::::::   |||   :::   :::
Db    10989 TACGAGACAGAGGGGGAAAGGAGA---------CAGCTAAAGATAGTATGTTACTGGGTT 11039

Qy      697 AspLys 698
               |||
Db    11040 AAGAAA 11045
```

RESULT 12

A81135

```
LOCUS       A81135                  2652 bp    DNA     linear   PAT 21-JAN-2000
DEFINITION  Sequence 1 from Patent EP0909818.
ACCESSION   A81135
VERSION     A81135.1  GI:6731609
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 2652)
  AUTHORS   Brown,J.R. and Wang,M.
  TITLE     Valyl tRNA synthetase (EC 6.1.1.9) from Streptococcus pneumoniae
  JOURNAL   Patent: EP 0909818-A 1 21-APR-1999;
            SMITHKLINE BEECHAM CORP (US)
```

```
FEATURES             Location/Qualifiers
     source          1..2652
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      733 a    612 c    621 g    686 t
ORIGIN

Alignment Scores:
Pred. No.:               5.67          Length:        2652
Score:                   130.00        Matches:       106
Percent Similarity:      32.81%        Conservative:  62
Best Local Similarity:   20.70%        Mismatches:    166
Query Match:             3.50%         Indels:        178
DB:                      6             Gaps:          25

US-10-008-355-2 (1-712) x A81135 (1-2652)

Qy    192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
          |||||||||                       |||      :::|||    |||
Db     76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 135

Qy    204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
           |||:::    ||||||                                      |||
Db    136 CCAAACGTTACAGGTAAACTT-------------------------------CACCTT 162

Qy    224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
           |||                                                 ::::::
Db    163 GGT-----------------------------------------------CACGCT 171

Qy    244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
              |||             :::  :::                  ||||||||:::
Db    172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAAACGCATGCAAGGTTTTGAT 228

Qy    264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                                  |||:::      ||||||          :::   |||
Db    229 -----------------ACCCTTTGGCTTCCTGGGATGGACCACGCAGGGATTGCCACT 270

Qy    283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                   ||||||:::|||:::      |||        |||   ::::::
Db    271 CAGGCTAAGGTAGAGGAGCGCTTGCGTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT 330

Qy    303 ------------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                        |||      ||||||:::           :::   ||||||
Db    331 GAGTCTTTCTTGACGAAAGTCTGGGAATGGAAAGAC-----------GAATATGCCACT 378

Qy    319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
              ||||||           |||        |||   :::|||    :::
Db    379 ACTATCAAGGAACAATGGGGCAAGATGGGGCTCTCTGTAGACTATTCTCGTGAGCGTTTC 438

Qy    338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
             ::::::   |||||||:::                        :::|||   :::
Db    439 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 474

Qy    358 AlaPheAlaAspTrpIleArgLys------------AsnGlyLysSerAlaValTyrGly 373
             |||   |||      :::|||              |||:::        :::
Db    475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 534

Qy    374 Asp---------------ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys 388
          |||               ||||||   :::|||      :::|||:::
Db    535 GACCCAGCAGCTCGCACAGCCCTTCTGATATTGAGGTGATTCACAAGGATGTGGAAGGT 594

Qy    389 AlaAsnArgGluMetThrTyrLeu---------------------------------- 396
          |||         |||   |||:::
Db    595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGTCCTTGAAGTTGCTACA 654

Qy    397 ------SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                |||||||:::||||||
Db    655 ACTCGTCCTGAGACCATGTTTGGGGACGTT------------------------------ 684

Qy    415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
          |||:::|||  ||||||:::                              |||   :::|||
Db    685 GCGGTTGCGGTCAACCCAGAA--------------------------GACCCGCGCTAC 717

Qy    435 LysAsp-----------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
          ||||||                  ||||||   :::      |||::::::|||
Db    718 AAGGACTTGATTGGTAAAAATGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy    449 MetLeuAsp-------------------------------------IleValArgArg 455
          :::   |||                                      |||
Db    775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy    456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
                |||   |||   |||               |||:::
Db    835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGTCATG 894

Qy    476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
             |||                |||   |||:::   |||   :::         |||
Db    895 AACGACGACGGAACCATGAATGACTTGGCCTTTGAATTTCAGGCATG---------GAC 945

Qy    496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
          :::|||   |||                  |||         ||||||      |||
Db    946 CGTTTTGAAGCTCGT--------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy    516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla------------ 531
                |||::::::   |||   |||       :::         :::
Db    991 GGTGCCCTCGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAGCGTACAGGT 1050

Qy    532 -----------------------------------------------------IleGlnAla 534
                                                               :::
Db   1051 GTTGTGGTTGAACCTCGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy    535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
          :::|||:::||||||            :::|||
Db   1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG--------------------------- 1143

Qy    555 ArgGluMetTyrProGlyArgAlaLeuProSer--------------------------- 565
             |||   ||||||   |||         :::
Db   1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203

Qy    566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
          |||         :::   ::::::   :::|||                     |||
Db   1204 GACTGGGTTATCTCTCGTCAGCTCTGGTGGGGTCAC-----------------CAAATC 1245

Qy    586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
             ||||||||||||            |||:::      :::   ::::::   |||:::
Db   1246 CCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy    606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
             |||   :::         |||::::::||||||
Db   1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335

RESULT 13
AR065782
LOCUS       AR065782                2652 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5849555.
ACCESSION   AR065782
VERSION     AR065782.1  GI:5995998
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2652)
  AUTHORS   Brown,J.R., Jaworski,D.D., Lawlor,E.J. and Wang,M.
  TITLE     Valyl trna polynucleotides from Streptococcus pneumoniae
  JOURNAL   Patent: US 5849555-A 1 15-DEC-1998;
FEATURES             Location/Qualifiers
     source          1..2652
                     /organism="unknown"
BASE COUNT      733 a    612 c    621 g    686 t
ORIGIN

Alignment Scores:
Pred. No.:               5.67          Length:        2652
Score:                   130.00        Matches:       106
Percent Similarity:      32.81%        Conservative:  62
Best Local Similarity:   20.70%        Mismatches:    166
```

```
Query Match:            3.50%      Indels:      178
DB:                     6          Gaps:        25

US-10-008-355-2 (1-712) x AR065782 (1-2652)

Qy    192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
          |||||||||                       |||       :::|||     |||
Db     76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 135

Qy    204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
          |||:::    ||||||                                      |||
Db    136 CCAAACGTTACAGGTAAACTT---------------------------------CACCTT 162

Qy    224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
          |||                                                   ::::::
Db    163 GGT---------------------------------------------------CACGCT 171

Qy    244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
             |||          :::   :::                  ||||||||||:::
Db    172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAAACGCATGCAAGGTTTTGAT 228

Qy    264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                            |||:::     ||||||                :::   |||
Db    229 ------------------ACCCTTTGGCTTCCTGGGATGGACCACGCAGGGATTGCCACT 270

Qy    283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                   ||||||:::|||:::        |||         |||    ::::::
Db    271 CAGGCTAAGGTAGAGGAGCGCTTGCGTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT 330

Qy    303 ------------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                         |||       ||||||:::                :::   ||||||
Db    331 GAGTCTTTCTTGACGAAAGTCTGGGAATGGAAAGAC------------GAATATGCCACT 378

Qy    319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
             ||||||            |||          |||    :::|||    :::
Db    379 ACTATCAAGGAACAATGGGGCAAGATGGGGCTCTCTGTAGACTATTCTCGTGAGCGTTTC 438

Qy    338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
             ::::::   ||||||:::                        :::|||       :::
Db    439 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 474

Qy    358 AlaPheAlaAspTrpIleArgLys-----------AsnGlyLysSerAlaValTyrGly 373
             |||   |||       :::|||               |||:::      :::
Db    475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 534

Qy    374 Asp---------------ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys 388
          |||                         ||||||   :::|||       :::|||:::
Db    535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTGGAAGGT 594

Qy    389 AlaAsnArgGluMetThrTyrLeu-------------------------------- 396
          |||          |||   |||:::
Db    595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGTCCTTGAAGTTGCTACA 654

Qy    397 ------SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                   ||||||:::||||||
Db    655 ACTCGTCCTGAGACCATGTTTGGGGACGTT----------------------------- 684

Qy    415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
          |||:::|||    ||||||:::                              |||    :::|||
Db    685 GCGGTTGCGGTCAACCCAGAA--------------------------GACCCGCGCTAC 717

Qy    435 LysAsp------------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
          ||||||                  ||||||    :::        |||:::::::|||
Db    718 AAGGACTTGATTGGTAAAAATGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy    449 MetLeuAsp---------------------------------------IleValArgArg 455
          :::   |||                                       |||
Db    775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy    456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
                |||   |||   |||            |||:::
Db    835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGTCATG 894

Qy    476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
             |||               |||       |||:::   |||    :::         |||
Db    895 AACGACGACGGAACCATGAATGACTTGGCCTTTGAATTTCAGGCATG---------GAC 945

Qy    496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
          :::|||   |||                         |||       ||||||     |||
Db    946 CGTTTTGAAGCTCGT---------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy    516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla------------ 531
                ||||:::::    |||    |||    :::               :::
Db    991 GGTGCCCTCGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAGCGTACAGGT 1050

Qy    532 ---------------------------------------------------IleGlnAla 534
                                                             :::
Db   1051 GTTGTGGTTGAACCTCGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy    535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
          :::|||:::||||||               :::|||
Db   1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG--------------------------- 1143

Qy    555 ArgGluMetTyrProGlyArgAlaLeuProSer--------------------------- 565
             |||   ||||||   |||          :::
Db   1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203

Qy    566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
          |||         :::    ::::::    :::|||                     |||
Db   1204 GACTGGGTTATCTCTCGTCAGCTCTGGTGGGGTCAC-------------------CAAATC 1245

Qy    586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
             ||||||||||||            |||:::       :::   :::::   |||:::
Db   1246 CCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy    606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
             |||   :::         |||:::::||||||
Db   1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335

RESULT 14
E26074
LOCUS       E26074                  2652 bp    DNA     linear   PAT 18-JUN-2001
DEFINITION  Novel valS.
ACCESSION   E26074
VERSION     E26074.1  GI:13025049
KEYWORDS    JP 1999113577-A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 2652)
  AUTHORS   James,R.B., Deborah,D.J., Elizabeth,J.L. and Ming,W.
  TITLE     Novel valS
  JOURNAL   Patent: JP 1999113577-A 1 27-APR-1999;
            SMITHKLINE BEECHAM CORP
COMMENT     OS   Unidentified
            PN   JP 1999113577-A/1
            PD   27-APR-1999
            PF   17-OCT-1997 JP 1997321886
            PR
            PI   JAMES R BURAUN,DEBORAH D JAWORUSUKI,ELIZABETH JANE LAWLOR, PI
             MING WANG
            PC   C12N15/09,A61K35/76,A61K38/43,A61K39/395,A61K48/00,C07K14/315,
            PC   C07K14/335,
            PC   C07K16/12,C12N5/10,C12N9/00,C12P21/02,C12Q1/00,G01N33/566, PC
             G01N33/569,
            PC   C12N15/00,A61K37/48,C12N5/00
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .2652
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .2652
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      733 a    612 c    621 g    686 t
```

```
ORIGIN

Alignment Scores:
Pred. No.:              5.67          Length:         2652
Score:                  130.00        Matches:        106
Percent Similarity:     32.81%        Conservative:   62
Best Local Similarity:  20.70%        Mismatches:     166
Query Match:            3.50%         Indels:         178
DB:                     6             Gaps:           25

US-10-008-355-2 (1-712) x E26074 (1-2652)

Qy    192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
          |||||||||                       |||       :::|||    |||
Db     76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 135

Qy    204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
          |||:::      ||||||                                  |||
Db    136 CCAAACGTTACAGGTAAACTT-------------------------------CACCTT 162

Qy    224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
          |||                                                :::::
Db    163 GGT-------------------------------------------------CACGCT 171

Qy    244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
             |||            :::   :::                 |||||||||:::
Db    172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAAACGCATGCAAGGTTTTGAT 228

Qy    264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                            |||:::      ||||||                :::   |||
Db    229 ------------------ACCCTTTGGCTTCCTGGGATGGACCACGCAGGGATTGCCACT 270

Qy    283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                   |||||:::|||:::       |||       |||    ::::::
Db    271 CAGGCTAAGGTAGAGGAGCGCTTGCGTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT 330

Qy    303 ------------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                      |||      ||||||:::                :::   ||||||
Db    331 GAGTCTTTCTTGACGAAAGTCTGGGAATGGAAAGAC------------GAATATGCCACT 378

Qy    319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
             ||||||            |||         |||   :::|||    :::
Db    379 ACTATCAAGGAACAATGGGGCAAGATGGGGCTCTCTGTAGACTATTCTCGTGAGCGTTTC 438

Qy    338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
             ::::::   ||||||:::                       :::|||      :::
Db    439 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 474

Qy    358 AlaPheAlaAspTrpIleArgLys------------AsnGlyLysSerAlaValTyrGly 373
             |||   |||     :::|||                |||:::       :::
Db    475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 534

Qy    374 Asp---------------ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys 388
          |||                  ||||||    :::|||       :::|||:::
Db    535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTGGAAGGT 594

Qy    389 AlaAsnArgGluMetThrTyrLeu------------------------------------ 396
          |||       |||   |||:::
Db    595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGTCCTTGAAGTTGCTACA 654

Qy    397 ------SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                      ||||||:::||||||
Db    655 ACTCGTCCTGAGACCATGTTTGGGGACGTT------------------------------ 684

Qy    415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
          |||:::|||   ||||||:::                           |||    :::|||
Db    685 GCGGTTGCGGTCAACCCAGAA---------------------------GACCCGCGCTAC 717

Qy    435 LysAsp------------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
          ||||||                  ||||||   :::      |||::::::|||
Db    718 AAGGACTTGATTGGTAAAAATGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy    449 MetLeuAsp---------------------------------------IleValArgArg 455
                                                         :::   |||
Db    775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy    456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
                |||   |||   |||                  |||:::
Db    835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGTCATG 894

Qy    476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
             |||                  |||     |||:::   |||   :::          |||
Db    895 AACGACGACGGAACCATGAATGACTTGGCCTTTGAATTTTCAGGCATG---------GAC 945

Qy    496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
          :::|||   |||                  |||         ||||||      |||
Db    946 CGTTTTGAAGCTCGT---------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy    516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla------------ 531
                   |||::::::   |||   |||   :::         :::
Db    991 GGTGCCCTCGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAGCGTACAGGT 1050

Qy    532 -----------------------------------------------------IleGlnAla 534
                                                            :::
Db   1051 GTTGTGGTTGAACCTCGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy    535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
          :::|||:::||||||               :::|||
Db   1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG--------------------------- 1143

Qy    555 ArgGluMetTyrProGlyArgAlaLeuProSer--------------------------- 565
             |||   ||||||   |||          :::
Db   1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203

Qy    566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
          |||         :::   ::::::   :::|||                        |||
Db   1204 GACTGGGTTATCTCTCGTCAGCTCTGGTGGGGTCAC------------------CAAATC 1245

Qy    586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
             ||||||||||||         |||:::      :::   ::::::   |||:::
Db   1246 CCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy    606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
             |||   :::         |||::::::||||||
Db   1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335
```

RESULT 15
AE013500

```
LOCUS       AE013500              9827 bp    DNA     linear   BCT 17-MAY-2002
DEFINITION  Methanosarcina mazei strain Goel, section 282 of 379 of the
            complete genome.
ACCESSION   AE013500 AE008384
VERSION     AE013500.1  GI:20907162
KEYWORDS    .
SOURCE      Methanosarcina mazei Goel.
  ORGANISM  Methanosarcina mazei Goel
            Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
            Methanosarcinaceae; Methanosarcina.
REFERENCE   1  (bases 1 to 9827)
  AUTHORS   Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,
            Martinez-Arias,R., Henne,A., Wiezer,A., Baeumer,S., Jacobi,C.,
            Brueggemann,H., Lienard,T., Christmann,A., Boemecke,M., Steckel,S.,
            Bhattacharyya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,
            Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.
  TITLE     The genome of Methanosarcina mazei: Evidence for lateral gene
            transfer between Bacteria and Archaea
  JOURNAL   J. Mol. Microbiol. Biotechnol. (2002) In press
REFERENCE   2  (bases 1 to 9827)
  AUTHORS   Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,
            Martinez-Arias,R., Henne,A., Wiezer,A., Baeumer,S., Jacobi,C.,
            Brueggemann,H., Lienard,T., Christmann,A., Boemecke,M., Steckel,S.,
            Bhattacharyya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,
            Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute
            for Microbiology and Genetics, Grisebachstrasse 8, Goettingen
```

```
           37077, Germany
FEATURES             Location/Qualifiers
     source          1. .9827
                     /organism="Methanosarcina mazei Goe1"
                     /strain="Goe1"
                     /db_xref="taxon:192952"
     gene            209. .1345
                     /gene="MM2559"
     CDS             209. .1345
                     /gene="MM2559"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAM32255.1"
                     /db_xref="GI:20907163"
                     /translation="MDEPIIEDVDDEDTNIAKKNKETNKETRKQLVTQKMDIFIPSLR
                     DMIKNEVINLSPRFQRRDRWNQNQQSKLIESVIMAVPIPPVFLAEEELGYYSVMDGKQ
                     RLTAIEEYLEDKYVLKGLKSWADLNGKKFSQLDPEAKGAITRRMISAIVILKESNLDI
                     KFDVFERINTGGEHLNFQEIRNCIHQGKLNDTLIELSENTFFRKCLNLPESKEELSKI
                     KIYQQMEDVQLVLRFFLLQDYKKMSGNLKSSLNNYMKNHKNITENELEAHRRLFNETI
                     EKVYGVYGNQSFRKWEPQNNLWSSRISAPFYDAVMYGFSKIDRSLIEGNKNKILDGTK
                     RLFEDSEFLDSIRKGTNSTEANNVRIEKFLSMLNSSVGLDESVL"
     gene            1329. .2069
                     /gene="MM2560"
     CDS             1329. .2069
                     /gene="MM2560"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAM32256.1"
                     /db_xref="GI:20907164"
                     /translation="MRASYDSFIGEISNLNSLLDRVEELGSSPQYLPIKAISSKSAII
                     MSSGLLEKYMKESFMEFIEQINEMNISEEYIDDKMWKTNRKNTLKALEYIDGKKLEAD
                     YEKVVFVYSESFRKNEKLNKPLLVKEAFSITNANPGPETVKNMFSNIGISKIFENTFF
                     QLEFGNEKFGDETRVTRTLKSFIELRNCIAHGYSGIPIPSIQEVQEYIKFLMKFVYTM
                     NEVLNQRLLIVQASHYKDCFKALDLFCK"
     gene            2412. .3509
                     /gene="MM2561"
     CDS             2412. .3509
                     /gene="MM2561"
                     /codon_start=1
                     /transl_table=11
                     /product="Conserved hypothetical protein"
                     /protein_id="AAM32257.1"
                     /db_xref="GI:20907165"
                     /translation="MNNHSEAESSGHGDNDTIPGWNEELELAFSAYKGSYVPGRVASR
                     HKTVCDVLLPAAVVQAGISGALQKIGKQPVVGDFVVVLDQPELGSRMIVNILPRKTCL
                     SRGAAGDGGGEQVIAANIDTIFIVTATGKDLNLRRLERYLAIVYSSGARPVILLNKID
                     LEDNPIRLVEKIQSIAGNVPVIPLSALSKTGLEALSPHLNPGETVALVGSSGVGKSTL
                     INSLLGETVQRTGGIREDDEKGRHTTTVRQMFLLPNGAVLIDNPGIREIQLGDSAEGI
                     EKAFSEIVDAASNCKFKDCTHRDEPGCAVLKAVKDGIIPEERLASYHRLTDELAFQSR
                     KSEIGLKRLEKERFKKIAVDIKKYKKSTGKL"
     gene            3666. .6239
                     /gene="MM2562"
     CDS             3666. .6239
                     /gene="MM2562"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAM32258.1"
                     /db_xref="GI:20907166"
                     /translation="MSRKLQDSAFRLLKKAAYDIDHDRYKKAFENLDKVEKIEKEVNN
                     PEISYHLLFFKGFAKYRIDELEESLALFGKALEISWDLFSKEPEKEDYRSFMRDTIGS
                     IGDILLKLEDPVKAREYISPMKGIFEKTVLGFEKLLESDPENPEYLEDFLKTIEHIGL
                     CFEAGELTGDAVPLFDKKFDLIEKLLKSETEISEHLSSLDDSLMNFGRMCEGEEYLEE
                     EYIEEAKRVYDRAIEIYGKVLDKDPENSETKVYLSYIYRYMAGMYSDQENLVKAEECY
                     KEALGLLESEIENEPENIPVSENIEVSENISVPKNILFSMAIADMYKDSGLIFSDTED
                     TTKARDYYTKARETFRGLINKYPDLLEADKNLASSLDELAELFSEIGDFESAEMCYKD
                     ELNIYENLLQKEPENPENPEYSLKLSDIYRELGDLFAEEEEIEKARAYYDKEIEIYEN
                     LHPAEADKLMLEANKADTWNQIGRLYAEEEPETAVQYHERALPIFLEAFESDPTSEFF
                     HEELLETLTVLGTLFKENREFEKAIRMYERVLEVQKQLLEIMPPASCHDHDHDHNLNH
                     SHNLNHSHNLNHTHDLNHDHDLNHDLDHDHEMDHDHALEHCHDLDIETTYFELGVLHS
                     EIGDEEKASVYHRQALERFEQIIAEQSGNPEILILTSGKAFFSGMSLLEKLESENKDS
                     SIARKYYELALRTIEKLYEAYPMNLKLQEMLASFAEQIGDIYRDADKLEETIQEYEYA
                     YRSLKILYENDPENPKHLHHMLGALNNIGIGYSVTDQQEKGKECFEKALALNESLAKS
                     NPEDLNPLRRNFLLFSNYANLLEELGDSKTAKEYRKKAEELDAKLTEEDPEWTSLVEE
                     L"
     gene            6716. .7825
                     /gene="MM2563"
     CDS             6716. .7825
                     /gene="MM2563"
                     /codon_start=1
                     /transl_table=11
                     /product="Ferredoxin"
                     /protein_id="AAM32259.1"
                     /db_xref="GI:20907167"
                     /translation="MASKVYFSDLRAKNPYESKISKIQKLFDKAGFVELLGAEDLTAI
                     KIHFGEYGNDGYINPVFVRQVVDKIRAAGAKPFITDTNTLYSGSRHNAVDHLTTAIEH
                     GFDYSVVRAPLIISDGLKSQNIAEVDIRQKHFKTVKIGSDIVEADSMIVMSHFKGHIM
                     AGFGGAIKNLAMGCAPAAGKKDQHYPTSPHVVEEKCIGCGNCVEICPVGAASLEGDVS
                     RINPGVCISCGQCMEVCPENAIDLNWEQDIPEFLECMTEYACGAVIGKENRVGYINFL
                     LKITPDCDCVPWSDAPIVPDIGILASTDPVALDQASYDLVNRQKGLVGSALHCNHEAG
                     ADKFKGTWPKIDGTHQLEYAELIGMGSREYELIEI"
     gene            complement(7993. .8355)
                     /gene="MM2564"
     CDS             complement(7993. .8355)
                     /gene="MM2564"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAM32260.1"
                     /db_xref="GI:20907168"
                     /translation="MKPKINSTSFGSITVRGKTFNYDILIRLDGKVEKRRKKLSKKLY
                     GTSHKISLEEAKYIHEKGAEKLIMGTGQTGYVELSKKAEKYFRKKDCKVKLLPTPEAI
                     KIWNKSKGKNIAMFHVTC"
     gene            8533. .9570
                     /gene="MM2565"
     CDS             8533. .9570
                     /gene="MM2565"
                     /EC_number="1.97.1.4"
                     /codon_start=1
                     /transl_table=11
                     /product="Pyruvate formate-lyase activating enzyme"
                     /protein_id="AAM32261.1"
                     /db_xref="GI:20907169"
                     /translation="MIHEAMFYEKLAENEVQCNLCAHRCRVRPGKRGICGVRENCEGS
                     LFSLVYGGVANEDVANIEQKPFYHYYPGSTAYSVGTIGCNFRCRHCQNWRLSRAAPED
                     AFLGTLSPSQLVMQAKMAGCQSVSWTYNEPTVWYEYTYEGARLAKEEGLGTSYVTNGY
                     ITPEALENIAPYLDAFSVDIKGFSEAFYREICGAKLAPVLETSRLAKELGIHVEVVNL
                     VIPTHNDSPDEIRELSRWVYENLGKDTPLHFNRFHPHYEMKHLPPTPINTLDRARSIA
                     IEEGMRFVYIGNVPGHPHENTYCPECGSLLIARGFFEVQEYNISPDKTCPKCGETIPI
                     TGEYRKSSGMI"
BASE COUNT     3165 a   1879 c   2132 g   2651 t
ORIGIN

Alignment Scores:
Pred. No.:              27.9           Length:         9827
Score:                  130.00         Matches:        145
Percent Similarity:     32.49%         Conservative:   126
Best Local Similarity:  17.39%         Mismatches:     279
Query Match:            3.50%          Indels:         284
DB:                     1              Gaps:           39

US-10-008-355-2 (1-712) x AE013500 (1-9827)

Qy       15 LeuLeuGlyAlaSerGlyValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeu 34
            |||:::|||:::|||||||||   |||:::               |||:::|||   |||
Db     3030 CTTGTAGGCTCTTCTGGTGTCGGAAAATCCACA------------CTTATCAACTCTCTT 3077

Qy       35 AsnGlnGluAsnLeuAspArg--------------------------------------- 41
                  |||   :::   |||
Db     3078 CTTGGTGAAACCGTCCAGAGAACCGGGGGTATCAGGGAAGACGATGAAAAAGGCAGGCAC 3137

Qy       42 ---------MetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAsp 58
                     :::|||::::::   |||   ||||||   :::                  |||
Db     3138 ACAACCACAGTCCGCCAGATG---TTTCTTCTTCCG---AATGGAGCAGTCCTTATAGAC 3191
```

```
Qy       59 LysProSerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrVal 78
              |||   |||              :::     |||         ||||||
Db     3192 AACCCCGGGATCAGGGAA------ATCCAGCTCGGGGATTCTGCCGAAGGGATTGAGAAA 3245

Qy       79 SerAspGlnGlyLeuIlePheThrAsnHisHisCysGlyTyr------------------ 92
            :::        ::::::            :::|||   :::
Db     3246 GCTTTTTCCGAAATCGTTGATGCAGCCAGCAATTGTAAGTTTAAAGACTGCACCCACCGG 3305

Qy       93 ---------------------------------------GlyAlaIleGlnSerGlnSer 99
                                                   |||   |||      :::
Db     3306 GACGAGCCGGGCTGTGCCGTACTGAAAGCTGTTAAGGATGGGATTATTCCTGAGGAACGG 3365

Qy      100 ThrValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeu 119
                :::      |||   |||        ::::::         |||:::
Db     3366 CTTGCCAGCTATCACAGGTTAACTGACGAACTTGCGTTTCAGTCAAGAAAGTCAGAAATC 3425

Qy      120 ProIleProGlyLeuSerValLysTyrLeuArgLysIle---------ValLysValThr 136
              :::      |||      :::      :::|||||||          :::|||
Db     3426 GGGTTGAAGAGGCTTGAAAAGGAAAGGTTCAAGAAAATAGCAGTGGACATTAAAAAATAT 3485

Qy      137 AspLysValGluGlyGlnLeu---------------------------LysGlyIleThr 147
             |||      |||:::|||                                 |||
Db     3486 AAGAAATCCACCGGGAAGCTGTAAGATAGGGTTTATTTTAAGAATGCAAGCATTATCAG 3545

Qy      148 AspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGluLeuAlaLysLysGlu 167
               :::    |||    |||       :::::::::
Db     3546 GGTTTTATCAATCGTTTTCGACCTGGGAGGAAACTG------------------------ 3581

Qy      168 AsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyrPhe 187
                                   :::  ||||||:::::        :::     |||
Db     3582 ------------GGAGGAGTAGTTTACATTGAACCCTCCCTGACTTACAAT--------- 3620

Qy      188 LeuIleValTyrAspValPheLysAspValArgMetValPheAlaProProSerSerVal 207
                                   :::    |||:::::::::::::::     ::::::
Db     3621 ------------------TACCCTGACCTACAATTATTATATGAAAAAACCTATGCTATT 3662

Qy      208 GlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSer 227
                                                 |||      |||    |||
Db     3663 TCT-----------------------------ATGTCAAGGAAGCTACAGGAT---TCT 3689

Qy      228 ValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLys 247
              ||||||:::                 :::    |||    |||      |||:::
Db     3690 GCTTTTCGTTTATTG--------------AAGAAAGCAGCCTACGATATAGATCATGAT 3734

Qy      248 ProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyr 267
              ||||||   :::
Db     3735 AGATATAAAAAAGCCTTT------------------------------------------ 3752

Qy      268 AlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGlu 287
                                    :::   |||:::
Db     3753 --------------------GAAAACCTTGATAAGGTT-------------------- 3770

Qy      288 AspArgIleGluAsnGlu---AsnAsnProArgIleGluValArgGlyIleLysGlnGly 306
            ::::::||||||    |||   |||||||||    |||
Db     3771 GAAAAGATCGAAAAGGAAGTGAATAACCCTGAGATTTCATAT------------------ 3812

Qy      307 IleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyr 326
                                                |||         :::||||||
Db     3813 --------------------------CATCTTCTGTTTTTAAAGGTTTTGCAAAATAC 3845

Qy      327 ---------------AlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArg 341
                             :::|||         :::      ::::::      ::::::
Db     3846 AGGATTGATGAACTCGAAGAGTCGCTTGCACTTTTCGGAAAGGCGCTGGAAATTAGCTGG 3905

Qy      342 GlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAsp 361
             |||                       |||      |||:::      :::
Db     3906 GACCTTTTT-----------------TCAAAAGAACCGGAAAAAGAAGATTACCGGTCT 3947

Qy      362 TrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLys 381
            ::::::|||         |||      ||||||:::|||      ||||||
Db     3948 TTTATGCGAGACACGATTGGAAGT------ATTGGAGATATCCTTCTGAAGCTTGAAGAT 4001

Qy      382 AlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPhe 401
             |||                            ||||||         :::   :::|||
Db     4002 CCCGTTAAAGCA--------------AGAGAGTATATCTCTCCGATGAAAGGGATTTTT 4046

Qy      402 GlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAsp 421
               |||   |||:::   |||   :::                  ||||||:::
Db     4047 GAAAAAACG---GTTCTGGGTTTTGAGAAGTTATTGGAGTCAGATCCCGAAAATCCGGAG 4103

Qy      422 --------------------------AlaHisAlaGlyIleLeuLysSerLeuAspAsp 432
                                       |||   |||:::                :::
Db     4104 TACCTGGAAGATTTCCTGAAAACGATCGAACACATAGGGCTCTGTTTTGAAGCCGGAGAA 4163

Qy      433 LysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIle 452
                   |||   :::|||      |||:::|||                  |||:::
Db     4164 TTGACCGGAGACGCAGTTCCACTTTTTGATAAAAAA--------------TTCGATCTC 4208

Qy      453 ValArgArgArgIleProAla--------------------------------------- 459
            :::   :::   :::   :::
Db     4209 ATTGAGAAATTGCTCAAAAGTGAAACTGAAATATCCGAACACCTCAGCAGCCTTGATGAT 4268

Qy      459 ------------------------------------------------------------ 459

Db     4269 TCTCTCATGAATTTTGGAAGGATGTGTGAGGGAGAGGAATATTTAGAAGAGGAGTATATA 4328

Qy      460 --------------------AspLysLeuProAspIlePheLysAsnValIleAspLys 472
                                |||:::      :::|||:::      |||:::||||||
Db     4329 GAAGAGGCAAAGCGCGTTTATGATCGGGCAATAGAGATTTACGGAAAAGTCCTCGATAAA 4388

Qy      473 ---LysPheLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValVal 491
                        :::||||||   |||      :::::::::
Db     4389 GACCCTGAAAACTCCGAAACCAAAGTTTACCTGAGTTACATATACCGTTACATGGCAGGT 4448

Qy      492 ProTyrSerAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAla--- 510
             |||||||||:::          ::::::|||:::   :::      |||   |||
Db     4449 ATGTATTCGGACCAG---GAAAATCTCGTAAAGGCAGAGGAATGTTATAAAGAAGCATTG 4505

Qy      511 -----------LysAlaIleGluLysAspProAla--------------------Val 519
                       ||||||   :::|||                               :::
Db     4506 GGCCTGCTTGAAAGCGAGATCGAAAATGAGCCTGAAAATATTCCGGTTTCTGAAAATATT 4565

Qy      520 GluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsn 539
            |||:::|||:::::::::                    :::   :::   |||:::|||:::
Db     4566 GAAGTTTCCGAAAACATATCAGTTCCCAAAAATATTCTTTTTTCTATGGCCATTGCAGAT 4625

Qy      540 AlaTyrAlaIleGluLysGly--------------------------------------- 546
              |||      :::   |||
Db     4626 ATGTACAAA---GATTCAGGCTTAATATTTTCAGATACGGAAGATACCACAAAAGCCAGG 4682

Qy      547 ---------------LysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArg 561
                          |||      |||   ||||||         ||||||
Db     4683 GACTACTATACGAAGGCGAGGGAAACTTTCCGGGGACTGATTAATAAATACCCGGAT--- 4739

Qy      562 AlaLeuProSerAspAlaAsnPheThrMetArgMetSer---------------TyrGly 576
             |||   :::|||   |||             :::
Db     4740 CTGCTTGAAGCCGATAAAAACCTTGCCAGCTCTTTAGATGAACTGGCAGAGCTGTTTTCA 4799

Qy      577 SerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLys 596
             |||      :::|||      :::      :::   :::               :::
Db     4800 GAGATCGGAGACTTCGAAAGTGCAGAAATGTGCTATAAGGACGAACTGAATATCTATGAA 4859

Qy      597 GlyValLeuGluLys------GlnAspProLysSerAspGluPheAlaValGlnGluAsn 614
               :::|||:::|||      ::::::|||::::::    |||::::::::::::
Db     4860 AACCTGTTGCAAAAAGAACCTGAAAACCCTGAAAACCCTGAATATTCTCTTAAG------ 4913

Qy      615 IleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHis 634
            :::   |||:::::::|||                              |||   |||   |||
Db     4914 CTTTCCGACATCTACCGG-----------------------GAGCTTGGTGACCTT--- 4946

Qy      635 IleAlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAsp 654
                  |||      :::|||                                 :::|||
Db     4947 ------TTTGCGGAAGAGGAAGAAATT---------GAAAAAGCAAGGGCTTATTATGAT 4991

Qy      655 LysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAsp 674
```

```
             |||                                                    :::
Db    4992 AAA----------------------------------------------------GAA 4997

Qy     675 IleGluPheGluProAspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPhe 694
           ||||||          :::|||           :::      :::
Db    4998 ATCGAAATTTACGAAAATCTTCACCCAGCAGAAGCCGATAAACTCATGTTAGAAGCTAAT 5057

Qy     695 MetIleAspLysTrpGlyGlnCysProArgLeuIleGlnGlu 708
                 |||   |||   |||      ||||||     |||
Db    5058 AAAGCCGATACCTGGAACCAGATTGGAAGGCTGTATGCAGAA 5099
```

Search completed: May 23, 2003, 11:33:15
Job time : 4235.8 secs

PD


```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         May 23, 2003, 07:32:23 ; Search time 316.182 Seconds
                                          (without alignments)
                                          5071.208 Million cell updates/sec

Title:          US-10-008-355-2
Perfect score:  3719
Sequence:       1 MQMKLKSILLGAALLLGASG..........LFMIDKWGQCPRLIQELKLI 712

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters:      4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110359_2525/app_query.fasta_1.1230
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_275_@runat_16052003_110359_2525 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      N_Geneseq_101002:*
                1:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                2:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                3:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                4:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                5:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                6:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                7:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                8:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                9:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                10:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                11:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                12:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                13:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                14:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                15:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                16:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                17:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                18:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                19:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
                20:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                21:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                22:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                23:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                24:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3719	100.0	2139	24	AAL43635	Porphyromonas ging
	2	146.5	3.9	4104	24	ABN68149	Streptococcus poly
	3	134.5	3.6	2853	20	AAX91754	Porphorymonas ging
	4	134.5	3.6	2859	20	AAX91621	Porphorymonas ging
	5	133	3.6	2652	23	AAS55647	Streptococcus pneu
	6	130	3.5	2652	20	AAX57027	S. pneumoniae valS
	7	130	3.5	2652	20	AAX32814	S. pneumoniae valy
	8	130	3.5	2652	21	AAA14369	DNA encoding Strep
	9	130	3.5	2652	24	ABK52414	DNA encoding Strep
c	10	129	3.5	6004	19	AAV52160	Streptococcus pneu
	11	127	3.4	2652	23	AAS55949	Streptococcus pneu
	12	126.5	3.4	4110	24	ABN68148	Streptococcus poly
	13	123.5	3.3	2142	20	AAZ31948	M. catarrhalis str
	14	122.5	3.3	4716	24	ABQ69595	Listeria innocua D
	15	122.5	3.3	4731	24	ABQ67873	Listeria innocua D
	16	122.5	3.3	495269	24	ABQ67195	Listeria innocua c
	17	121.5	3.3	2649	24	ABN66404	Streptococcus poly
	18	121.5	3.3	2155561	24	ABN71527	Streptococcus poly
	19	121	3.3	2646	24	ABN66405	Streptococcus poly
	20	119.5	3.2	3271	22	AAI64839	XRN-100 coding seq
	21	117.5	3.2	1664976	19	AAV21209	Methanococcus jann
c	22	117	3.1	1830121	17	AAT42063	Haemophilus influe
	23	116.5	3.1	15424	21	AAC81763	Porcine Lelystad v
	24	116.5	3.1	15424	21	AAC81764	Porcine Lelystad v
	25	116	3.1	3598	21	AAC77776	Human cancer assoc
	26	115.5	3.1	6822	24	ABN71008	Streptococcus poly
	27	115.5	3.1	6834	24	ABN68971	Streptococcus poly
	28	115	3.1	4853	17	AAT13139	Alpha-D-glucosyltr
	29	114.5	3.1	1974	24	AAL46803	H influenzae BVH-N
	30	114	3.1	2136	17	AAT11244	Neisseria meningit
	31	114	3.1	3420	22	AAH14452	Human cDNA sequenc
	32	114	3.1	3431	21	AAC76951	Human ORFX ORF2506
	33	113.5	3.1	2856	23	AAS52731	E. coli DNA for ce
	34	113	3.0	8930	19	AAV22834	Haemophilus paraga
	35	112.5	3.0	36471	21	AAA81453	N. meningitidis pa
	36	112.5	3.0	349980	21	AAF21611	Neisseria meningit
	37	112.5	3.0	837096	21	AAA81489	N. meningitidis pa
	38	112	3.0	2175	23	AAS79571	DNA encoding novel
	39	112	3.0	2175	23	AAS81718	DNA encoding novel
c	40	112	3.0	3285	22	AAH54398	S. epidermidis gen
	41	112	3.0	56506	21	AAA69168	Bacteriophage Dp-1
c	42	111	3.0	4411529	22	AAI99682	Mycobacterium tube
	43	110.5	3.0	3279	14	AAQ51556	Sequence encoding
	44	110.5	3.0	12665	19	AAV52267	Streptococcus pneu
	45	110.5	3.0	15450	21	AAA27809	North American por

ALIGNMENTS

```
RESULT 1
AAL43635
ID   AAL43635 standard; DNA; 2139 BP.
XX
AC   AAL43635;
XX
DT   05-SEP-2002  (first entry)
XX
DE   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) coding sequence.
XX
KW   Dipeptidylpeptidase-7; DPP-7; gene; ds; enzyme; amidolytic cleavage;
KW   DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW   periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2139
```

```
FT                   /*tag= a
FT                   /product= "Porphyromonas gingivalis DPP-7"
XX
PN   WO200238742-A2.
XX
PD   16-MAY-2002.
XX
PF   08-NOV-2001; 2001WO-US46782.
XX
PR   08-NOV-2000; 2000US-246827P.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J,  Potempa JS,  Banbula A,  Bugno M;
XX
DR   WPI; 2002-490075/52.
DR   P-PSDB; AAO15205.
XX
PT   Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT   the dipeptidylpeptidase for protecting an animal from periodontal
PT.  disease caused by Porphyromonas gingivalis  -
XX
PS   Claim 11; Fig 4; 65pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC   enzyme has amidolytic activity for cleavage of a peptide bond between the
CC   second and third amino acids from the N-terminal end of a target peptide.
CC   The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC   substituent on the alpha-carbon atom of the second amino acid from the
CC   N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC   useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC   reducing the growth of a bacterium and protecting an animal from a
CC   periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC   or periodontitis). The present DNA sequence encodes the Porphyromonas
CC   gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the invention.
XX
SQ   Sequence 2139 BP; 544 A; 543 C; 565 G; 487 T; 0 other;

Alignment Scores:
Pred. No.:               0              Length:         2139
Score:                   3719.00        Matches:        712
Percent Similarity:      100.00%        Conservative:   0
Best Local Similarity:   100.00%        Mismatches:     0
Query Match:             100.00%        Indels:         0
DB:                      24             Gaps:           0

US-10-008-355-2 (1-712) x AAL43635 (1-2139)

Qy       1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60

Qy      21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAsp 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120

Qy      41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180

Qy      61 SerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240

Qy      81 GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300

Qy     101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuPro 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360

Qy     121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420

Qy     141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480

Qy     161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540

Qy     181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600

Qy     201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660

Qy     221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720

Qy     241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780

Qy     261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840

Qy     281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900

Qy     301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960

Qy     321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020

Qy     341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080

Qy     361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140

Qy     381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200

Qy     401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260

Qy     421 AspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSer 440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320

Qy     441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAsp 460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380

Qy     461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440

Qy     481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500

Qy      501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560

Qy      521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620

Qy      541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680

Qy      561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740

Qy      581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800

Qy      601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860

Qy      621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920

Qy      641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980

Qy      661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040

Qy      681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGly 700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100

Qy      701 GlnCysProArgLeuIleGlnGluLeuLysLeuIle 712
            ||||||||||||||||||||||||||||||||||||
Db     2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATC 2136
```

RESULT 2
ABN68149

```
ID   ABN68149 standard; DNA; 4104 BP.
XX
AC   ABN68149;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 4211.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus pyogenes.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB04789.
XX
PR   27-OCT-2000; 2000GB-0026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J,  Masignani V,  Margarit Ros YI,  Grandi G,  Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP27518.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein -
XX
PS   Claim 7; Page 3578; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence 4104 BP; 1458 A; 619 C; 821 G; 1206 T; 0 other;

Alignment Scores:
Pred. No.:              0.00104       Length:        4104
Score:                  146.50        Matches:       172
Percent Similarity:     31.98%        Conservative:  103
Best Local Similarity:  20.00%        Mismatches:    223
Query Match:            3.94%         Indels:        363
DB:                     24            Gaps:          44

US-10-008-355-2 (1-712) x ABN68149 (1-4104)

Qy       35 AsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeu 54
            |||:::|||||    |||    :::   ::::::      ||||||   |||
Db     1825 AATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAA--- 1881

Qy       55 TyrSerPheAspLysProSerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThr 74
                         |||:::       |||           :::        :::
Db     1882 ---------GATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCTCAC--------- 1923

Qy       75 GlyIleThrValSerAspGlnGlyLeu------IlePheThrAsnHisHisCysGlyTyr 92
                                 |||:::    :::       :::            :::   |||:::
Db     1924 ------CTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGG 1977

Qy       93 GlyAlaIle--------------------------------GlnSerGlnSerThrVal 101
            |||   :::                                       ||||||    |||:::
Db     1978 GGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATA 2037

Qy      102 AspHisAspTyrLeuArg---AspGlyPheValSerArg-------------------- 113
                  |||:::|||:::    |||||||||   :::|||
Db     2038 ---TTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCAT 2094

Qy      114 -----------ThrMetGlyGluGluLeuProIleProGlyLeuSer------------ 125
                        |||        |||::::::                 :::|||
Db     2095 GATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGAT 2154

Qy      126 ----------------------------------------------ValLysTyrLeuArg 130
                                                          |||    :::
Db     2155 AGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTA 2214
```

```
Qy      131 LysIleValLysValThrAsp---------LysValGluGlyGln--------------- 142
            :::   |||||||||  |||         ||||||  |||:::
Db     2215 CAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGCGGCATAAGCCAGAAAAT 2274

Qy      143 ------------------------------------LeuLysGlyIleThrAspGluMet 150
                                                |||||||      :::
Db     2275 ATCGTTATTGAAATGGCACGTGAAAATCACACAACTCAAAAGGGCCAGAAAAATTCGCGA 2334

Qy      151 GluArgLeuArgLysAlaGlnGluValCysGlnGluLeuAlaLys---------LysGlu 167
            |||||:::::::::   :::|||     :::||||||           ||||||
Db     2335 GAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAG 2394

Qy      168 AsnAlaAspGluAsn---GlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyr 186
            :::     ||||||  ||||||                          ||||||
Db     2395 CATCCTGTTGAAAATACTCAATTG------------------------CAAAATGAAAAG 2430

Qy      187 PheLeuIleValTyr---------------AspValPheLysAspValArgMetValPhe 201
                :::   |||                 |||::::::   |||        :::
Db     2431 CTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATT 2490

Qy      202 AlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArg 221

Db     2490 ------------------------------------------------------------ 2490

Qy      222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGlu 241
                                                         ||||||   ::::::
Db     2491 ---------------------------------------------AATCGTTTAAGTGAT 2505

Qy      242 TyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGly 261
            |||      |||:::                               :::    |||
Db     2506 TATGATGTCGATCAC------------------------------ATTGTTCCACAAAGT 2535

Qy      262 TyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu 281
            :::       ||||||                           |||   |||      :::
Db     2536 TTCCTTAAAGACGAT---------------------------TCAATAGACAATAAGGTC 2568

Qy      282 ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArg 301
              :::         :::::::|||   :::::::::|||   |||      ||||||
Db     2569 TTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTC 2628

Qy      302 GlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLys 321
                     :::         |||::::::   :::::::|||         |||:::   |||
Db     2629 AAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAG 2688

Qy      322 TyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArg 341
            :::                           |||      |||         |||
Db     2689 TTT---------------------GATAATTTAACGAAGCTGAACGTGGA---------- 2718

Qy      342 GlyLeuAlaArgLeuAspValIleGly---------ArgLysArgAlaGluGluArgAla 358
            ||||||:::   ||||||      |||         |||:::      |||   |||
Db     2719 GGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAA 2778

Qy      359 PheAlaAspTrpIle------------------ArgLysAsnGlyLys------------ 368
                     :::                  |||   |||   |||
Db     2779 ATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAAT 2838

Qy      368 ------------------------------------------------------------ 368

Db     2839 GATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTC 2898

Qy      368 ------------------------------------------------------------ 368

Db     2899 CGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGAT 2958

Qy      368 ------------------------------------------------------------ 368

Db     2959 GCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCG 3018

Qy      369 SerAlaValTyrGly---------------AspValLeuSerSerLeuGluLysAlaTyr 383
                  |||||||||               ||||||   :::   |||:::
Db     3019 GAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAG 3078

Qy      384 LysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu---PheGly 402
            :::|||   ||||||   :::      :::      |||      :::   |||
Db     3079 CAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTC 3138

Qy      403 GlyThrGluVal----ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAs 421
               ||||||:::   ::::::   :::|||||||||||||||                ::
Db     3139 AAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTA-ATCGAAACTAA 3197

Qy      421 pAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLe 441
            :       |||   :::   :::   ||||||   :::|||:::            :::
Db     3198 TGGGGAAACTGGA------GAAATTGTCTGGGATAAAGGGCGAGATTTT------GCCAC 3245

Qy      441 uAspArgLysValLeuPro---AlaMetLeuAspIleValArgArgArg----------- 456
             ||||||||||||         ::::::|||||||||::::::
Db     3246 AGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGAC 3305

Qy      457 ------------------------IlePro----------AlaAspLysLeuProAspIl 465
                                    :::|||           :::|||||||||
Db     3306 AGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAATTCGGACAAGCTT-------- 3357

Qy      465 ePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPheVa 485
                     |||   :::|||         |||   |||||||||
Db     3358 -------------ATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGT----- 3399

Qy      485 lPheAspLysSerValValProTyrSerAspLysPheHisAla----------------- 499
             ||||||      |||   ||||||         |||
Db     3400 -TTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAA 3458

Qy      500 -------MetLeuLysSer---------------------------MetAspLysGluLy 508
                   |||||||||                           |||::::::
Db     3459 ATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTC 3518

Qy      508 sPheAla------------------------------------LysAlaIleGluLysAs 516
             |||                                        |||   :::::|||||
Db     3519 CTTTGAAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGA 3578

Qy      516 pProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAl 536
            |     ::::::|||   |||
Db     3579 CTTAATCATTAAACTACCTAAA-------------------------------------- 3600

Qy      536 aMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGl 556
                :::   :::   :::|||   |||::::::         |||         ||
Db     3601 ----TATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGA 3656

Qy      556 uMetTyrProGlyArg------AlaLeuProSerAsp---AlaAsnPheThrMetArgMe 573
            |:::   |||         ||||||||||||         ||||||
Db     3657 ATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGC 3716

Qy      573 tSer---TyrGlySerIleLysGlyTyrGluProGlnAspGly------------AlaTr 588
             |||   |||         :::||||||   |||:::|||                ::
Db     3717 TAGTCATTATGAAAAGTTGAAGGGT---AGTCCAGAAGATAACGAACAAAAACAATTGTT 3773

Qy      588 pTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGl 608
            :         |||      |||   |||:::::::         :::         ||
Db     3774 TGTGGAGCAGCAT---------AAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGA 3824

Qy      608 uPheAlaValGln----------------------------------------------- 612
            ||||:::   :::
Db     3825 ATTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAA 3884

Qy      613 ----------------------------------GluAsnIleLeuAspLeuPheArgTh 621
                                              |||||||||:::   ||||||
Db     3885 CAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTT 3944

Qy      621 rLysAsnTyrGly---------------------------------------------Ar 626
               |||   |||                                              ||
Db     3945 GACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACG 4004

Qy      626 gTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly 645
            ||||            ::::::   |||   |||   :::      |||||||||
Db     4005 ATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGT 4062

RESULT 3
AAX91754
```

```
ID   AAX91754 standard; DNA; 2853 BP.
XX
AC   AAX91754;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Porphorymonas gingivalis protein PG67 encoding DNA.
XX
KW   Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW   vaccine; antigenic; ds.
XX
OS   Porphorymonas gingivalis.
XX
PN   WO9929870-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-AU01023.
XX
PR   04-AUG-1998;   98AU-0005028.
PR   10-DEC-1997;   97AU-0000839.
PR   31-DEC-1997;   97AU-0001182.
PR   30-JAN-1998;   98AU-0001546.
PR   10-MAR-1998;   98AU-0002264.
PR   09-APR-1998;   98AU-0002911.
PR   23-APR-1998;   98AU-0003128.
PR   05-MAY-1998;   98AU-0003338.
PR   22-MAY-1998;   98AU-0003654.
PR   29-JUL-1998;   98AU-0004917.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Agius CT,  Barr IG,  Hocking DM,  Margetts MB,  Patterson MA;
PI   Ross BC,  Rothel LJ,  Webb EA;
XX
DR   WPI; 1999-385613/32.
DR   P-PSDB; AAY34536.
XX
PT   Antigenic Porphorymonas gingivalis peptides for preventing
PT   gingivitis
XX
PS   Claim 12; Page 244-245; 588pp; English.
XX
CC   AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC   Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC   AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC   isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC   activity with a vaccine mechanism of action. The PG polypeptides can be
CC   used as vaccines especially against Porphorymonas gingivalis. Probes can
CC   be used to detect Porphorymonas gingivalis in standard hybridisation
CC   assays. Porphorymonas gingivalis is involved in periodontal disease
CC   especially gingivitis.
XX
SQ   Sequence 2853 BP; 736 A; 685 C; 691 G; 741 T; 0 other;

Alignment Scores:
Pred. No.:              0.00909      Length:         2853
Score:                  134.50       Matches:        105
Percent Similarity:     34.41%       Conservative:   66
Best Local Similarity:  21.13%       Mismatches:     202
Query Match:            3.62%        Indels:         124
DB:                     20           Gaps:           23

US-10-008-355-2 (1-712) x AAX91754 (1-2853)

Qy    221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArg-ProAl 240
          ||||||         |||   :::|||  ||||||  ||||||      |||
Db    717 CGTCACAGGTCCGGTCAATCTGGTTATCGCCGATATGCCGCTGCCGATAGGTCTTCCTTT 776

Qy    240 aGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGl 260
              |||          |||:::      |||  :::   :::   |||
Db    777 CGGCTATTTT-----------CCCTTTTCCAACAAATACTCTTCCGGTATATTGATGCC 824

Qy    260 nGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTy 280
            |||      |||:::           ||||||                   ||
Db    825 CACGTACGGAGAGGACAAT---------CGCTATGGATTT-----------------TA 857

Qy    280 rLeuThrSerTrpGly-----------ValGluAspArgIleGluAsnGluAsnAsnPr 296
          ||||  :::  |||                  |||  |||:::
Db    858 TTTGAGGAATGGTGGATATTATTTGCCTTCAGCGACTATATCGAT------------- 903

Qy    296 oArgIleGluValArgGly------IleLysGlnGlyIleTrpLysGluAlaMetSerAl 314
             :::   :::||||||          :::|||   |||      :::      ::
Db    904 ----TTGGCATTGCGTGGGGAGATCTTTTCCAAAGGGTCATGGGGCATTTCAGCCCAATC 959

Qy    314 aAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLy 334
          :              |||   ||||||         :::        |||||||||
Db    960 GAAATATAAGAAGAGGTATAAGTACAACGGCTCGTTC------GAAGCCAATTAT----- 1008

Qy    334 sAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAl 354

Db   1008 ------------------------------------------------------------ 1008

Qy    354 aGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAs 374
                                :::    |||:::|||            |||   |||||
Db   1009 ---------------------CTGGTATCGAAGTCCGGCGACAAATACGTGCCCGGAGA 1046

Qy    374 pValLeuSer-----------SerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAs 390
          |                    ::::::        :::   :::      |||||||
Db   1047 CTACAGCAAGACCACCAGTCTGAATATCCGATGGACACACAGTCAGGATCCGAAGGCCAA 1106

Qy    390 nArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAl 410
          |                    :::||||||                           ::
Db   1107 TCCT--------------TTGCAAACGTTGTCGGCCAATGTCAATTTTGCCACCGGGAG 1151

Qy    410 aGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla------GlyIleLeuLy 428
          :  |||   |||:::|||   |||      |||   :::|||               ::
Db   1152 CTATTTCCAGAATTCGCTGAATACCACCTATGATGTCAATGCCCGTACTGCTACGACACG 1211

Qy    428 sSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAl 448
          :|||            |||         |||                  :::      ::
Db   1212 AAGTTCGGCCGTGAGCTATTCGCGCAAGTTTCCGGGTACTCCTTTTTCGATTACGGGTAG 1271

Qy    448 aMetLeuAspIle---------ValArgArgArgIleProAlaAspLysLeuProAspIl 465
          :|||   ||||||          :::|||            :::      |||||:::::
Db   1272 CATG---GATATCAGCCAGAACATGCGCGATACGACGGTGAGCCTTACCTTGCCGAATCT 1328

Qy    465 ePheLysAsnValIleAspLys---LysPheLysGlyAspThrLysLysTyrAlaAspPh 484
          :      |||:::      :::      ||||||      |||:::         :::
Db   1329 TTCGATTAATATGTCCACGCGTTATCCTTTCAAGCGGAAGACCCGTGTAGGACCGGAGCG 1388

Qy    484 eValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLysSerMe 504
             :::::|||         |||   ||||||   :::            :::   :::
Db   1389 ATGGTACGAGAAGTTGAGTGTGGGCTATTCCGGTCAGCTTCGCAATAGTATCTTGACAAA 1448

Qy    504 tAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLysSe 524
           :::|||:::                    :::   :::|||            :::   ||
Db   1449 AGAGAAAGATTTGCTCCAGAGCAATCTCGTGCGCGATTGGAAGAATGGTATGCGTCATTC 1508

Qy    524 rValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGl 544
           ||||     :::                    |||   :::               :::
Db   1509 CGTACCGATCAGTTTGACTGTCCCTTTGTTGGATTATATCAATCTGACTATGGGGGTTAA 1568

Qy    544 uLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyr---------ProGlyAr 561
                :::      :::      |||:::|||:::   :::
Db   1569 CTACAATGAGTGGTGGTACACGAAAGGCATACGGAAGTCGTGGAATGAGGATAAGAAAAC 1628

Qy    561 gAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTy 581
             ||||||||||||         :::      |||      |||               ||
Db   1629 ATTCCTGCCTTCGGACACGACCTATAAATTCCGCAGACTGTAC------------GATTA 1676

Qy    581 rGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLy 601
          |            |||               ||||||      |||:::         ::
Db   1677 CAGTCTGTCGGCAGGC------------TTATCTACCACATTGTAGGGTATGTTCAAGCC 1724

Qy    601 sGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu-------- 618
```

```
                 |||            |||:::          |||:::::     :::
Db   1725 TTGGAAACCT-----------TTTTCCTTCGGAGGCAATCTCATTATGATCCGTCATCG 1772

Qy    619 -------------------------------------PheArgThrLysAsnTyrGlyAr 626
                                               |||      :::   ||||||
Db   1773 CTTCACGCCCACTGTCAGTTTCTCCTATATGCCGGACTTCACGAAACGCCGATATGGCTT 1832

Qy    626 gTyrAla------------------GluAsnGlyGlnLeuHisIleAlaPheLeuSerAs 640
           :::                      :::||||||:::||||||           |||
Db   1833 TTGGGAGCTTCTTGAGCATACGGATCAGAACGGCAAGCTGCATACGCTGCTCTACTCTCC 1892

Qy    640 n------AsnAspIleThrGly------------GlyAsnSerGlySerProValPheAs 654
                        |||  |||            ||||||:::||||||
Db   1893 TTATTTCGAGCAGATATTCGGTGCTCCCTCCATGGGCAATGCAGGATCT----------- 1941

Qy    654 pLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAla 670
                          :::      :::||||||   |||   ||||||
Db   1942 ----------------GTCAATTTCTCTTTTGACAACAACTTAGAGGCC 1974
```

```
RESULT 4
AAX91621
ID   AAX91621 standard; DNA; 2859 BP.
XX
AC   AAX91621;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Porphorymonas gingivalis protein PG67 ORF encoding DNA.
XX
KW   Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW   vaccine; antigenic; ds.
XX
OS   Porphorymonas gingivalis.
XX
PN   WO9929870-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-AU01023.
XX
PR   04-AUG-1998;   98AU-0005028.
PR   10-DEC-1997;   97AU-0000839.
PR   31-DEC-1997;   97AU-0001182.
PR   30-JAN-1998;   98AU-0001546.
PR   10-MAR-1998;   98AU-0002264.
PR   09-APR-1998;   98AU-0002911.
PR   23-APR-1998;   98AU-0003128.
PR   05-MAY-1998;   98AU-0003338.
PR   22-MAY-1998;   98AU-0003654.
PR   29-JUL-1998;   98AU-0004917.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Agius CT,  Barr IG,  Hocking DM,  Margetts MB,  Patterson MA;
PI   Ross BC,  Rothel LJ,  Webb EA;
XX
DR   WPI; 1999-385613/32.
DR   P-PSDB; AAY34403.
XX
PT   Antigenic Porphorymonas gingivalis peptides for preventing
PT   gingivitis
XX
PS   Claim 12; Page 153-154; 588pp; English.
XX
CC   AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC   Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC   AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC   isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC   activity with a vaccine mechanism of action. The PG polypeptides can be
CC   used as vaccines especially against Porphorymonas gingivalis. Probes can
CC   be used to detect Porphorymonas gingivalis in standard hybridisation
CC   assays. Porphorymonas gingivalis is involved in periodontal disease
CC   especially gingivitis.
XX
SQ   Sequence 2859 BP; 738 A; 687 C; 692 G; 742 T; 0 other;

Alignment Scores:
Pred. No.:                0.00912        Length:         2859
Score:                    134.50         Matches:        105
Percent Similarity:       34.41%         Conservative:   66
Best Local Similarity:    21.13%         Mismatches:     202
Query Match:              3.62%          Indels:         124
DB:                       20             Gaps:           23

US-10-008-355-2 (1-712) x AAX91621 (1-2859)
```

```
Qy    221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArg-ProAl 240
          ||||||            |||   :::|||   ||||||   ||||||       |||
Db    723 CGTCACAGGTCCGGTCAATCTGGTTATCGCCGATATGCCGCTGCCGATAGGTCTTCCTTT 782

Qy    240 aGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGl 260
            |||        |||:::                   |||  :::   :::   |||
Db    783 CGGCTATTTT------------CCCTTTTCCAACAAATACTCTTCCGGTATATTGATGCC 830

Qy    260 nGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTy 280
              |||      |||:::              ||||||                   ||
Db    831 CACGTACGGAGAGGACAAT---------CGCTATGGATTT------------------TA 863

Qy    280 rLeuThrSerTrpGly------------ValGluAspArgIleGluAsnGluAsnAsnPr 296
          ||||   :::   |||                  |||   |||:::
Db    864 TTTGAGGAATGGTGGATATTATTTTGCCTTCAGCGACTATATCGAT-------------- 909

Qy    296 oArgIleGluValArgGly------IleLysGlnGlyIleTrpLysGluAlaMetSerAl 314
              :::   :::||||||            :::|||   |||      :::      ::
Db    910 ----TTGGCATTGCGTGGGGAGATCTTTTCCAAAGGGTCATGGGGCATTTCAGCCCAATC 965

Qy    314 aAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLy 334
          :            |||  ||||||        :::            |||||||||
Db    966 GAAATATAAGAAGAGGTATAAGTACAACGGCTCGTTC------GAAGCCAATTAT----- 1014

Qy    334 sAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAl 354

Db   1014 ------------------------------------------------------------ 1014

Qy    354 aGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAs 374
                                   :::   |||:::|||         |||   |||||
Db   1015 ----------------------CTGGTATCGAAGTCCGGCGACAAATACGTGCCCGGAGA 1052

Qy    374 pValLeuSer------------SerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAs 390
          |                     ::::::         :::   :::      |||||||
Db   1053 CTACAGCAAGACCACCAGTCTGAATATCCGATGGACACACAGTCAGGATCCGAAGGCCAA 1112

Qy    390 nArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAl 410
          |                     :::||||||                           ::
Db   1113 TCCT---------------TTGCAAACGTTGTCGGCCAATGTCAATTTTGCCACCGGGAG 1157

Qy    410 aGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla------GlyIleLeuLy 428
          :   |||   |||:::|||   |||      |||   :::|||               ::
Db   1158 CTATTTCCAGAATTCGCTGAATACCACCTATGATGTCAATGCCCGTACTGCTACGACACG 1217

Qy    428 sSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAl 448
          :|||            |||            |||               :::      ::
Db   1218 AAGTTCGGCCGTGAGCTATTCGCGCAAGTTTCCGGGTACTCCTTTTTCGATTACGGGTAG 1277

Qy    448 aMetLeuAspIle---------ValArgArgArgIleProAlaAspLysLeuProAspIl 465
          :|||   ||||||         :::|||            :::      ||||||::::
Db   1278 CATG---GATATCAGCCAGAACATGCGCGATACGACGGTGAGCCTTACCTTGCCGAATCT 1334

Qy    465 ePheLysAsnValIleAspLys---LysPheLysGlyAspThrLysLysTyrAlaAspPh 484
          :    |||:::     :::         ||||||      |||:::         :::
Db   1335 TTCGATTAATATGTCCACGCGTTATCCTTTCAAGCGGAAGACCCGTGTAGGACCGGAGCG 1394

Qy    484 eValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLysSerMe 504
              :::::|||       |||   ||||||   :::            :::   :::
Db   1395 ATGGTACGAGAAGTTGAGTGTGGGCTATTCCGGTCAGCTTCGCAATAGTATCTTGACAAA 1454
```

```
Qy     504 tAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLysSe 524
              :::|||:::                :::   :::|||          :::      ||
Db    1455 AGAGAAAGATTTGCTCCAGAGCAATCTCGTGCGCGATTGGAAGAATGGTATGCGTCATTC 1514

Qy     524 rValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGl 544
            ||||        :::                   |||    :::              :::
Db    1515 CGTACCGATCAGTTTGACTGTCCCTTTGTTGGATTATATCAATCTGACTATGGGGGTTAA 1574

Qy     544 uLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyr---------ProGlyAr 561
                 :::      :::      |||:::|||:::   :::
Db    1575 CTACAATGAGTGGTGGTACACGAAAGGCATACGGAAGTCGTGGAATGAGGATAAGAAAAC 1634

Qy     561 gAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTy 581
               |||||||||||         :::      |||       |||            ||
Db    1635 ATTCCTGCCTTCGGACACGACCTATAAATTCCGCAGACTGTAC-----------GATTA 1682

Qy     581 rGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLy 601
            |          |||                  ||||||      |||:::     :::
Db    1683 CAGTCTGTCGGCAGGC-----------TTATCTACCACATTGTACGGTATGTTCAAGCC 1730

Qy     601 sGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu-------- 618
                 |||          |||:::          |||::::::    :::
Db    1731 TTGGAAACCT-----------TTTTCCTTCGGAGGCAATCTCATTATGATCCGTCATCG 1778

Qy     619 ------------------------------------PheArgThrLysAsnTyrGlyAr 626
                                               |||      :::   ||||||
Db    1779 CTTCACGCCCACTGTCAGTTTCTCCTATATGCCGGACTTCACGAAACGCCGATATGGCTT 1838

Qy     626 gTyrAla-----------------GluAsnGlyGlnLeuHisIleAlaPheLeuSerAs 640
             :::                  :::||||||:::||||||           |||
Db    1839 TTGGGAGCTTCTTGAGCATACGGATCAGAACGGCAAGCTGCATACGCTGCTCTACTCTCC 1898

Qy     640 n------AsnAspIleThrGly-----------GlyAsnSerGlySerProValPheAs 654
                     |||   |||                |||||:::||||||
Db    1899 TTATTTCGAGCAGATATTCGGTGCTCCCTCCATGGGCAATGCAGGATCT----------- 1947

Qy     654 pLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAla 670
                             :::      :::||||||   |||   ||||||
Db    1948 ---------------GTCAATTTCTCTTTTGACAACAACTTAGAGGCC 1980
```

```
RESULT 5
AAS55647
ID   AAS55647 standard; DNA; 2652 BP.
XX
AC   AAS55647;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Streptococcus pneumoniae DNA for cellular proliferation protein #218.
XX
KW   Antisense; ds; prokaryotic cellular proliferation gene;
KW   antibiotic; antibacterial; drug design.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   P-PSDB; AAU37788.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 27; Seq ID No 9284; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence encodes an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2652 BP; 736 A; 609 C; 618 G; 689 T; 0 other;

Alignment Scores:
Pred. No.:              0.0115        Length:         2652
Score:                  133.00        Matches:        106
Percent Similarity:     33.20%        Conservative:   64
Best Local Similarity:  20.70%        Mismatches:     164
Query Match:            3.58%         Indels:         178
DB:                     23            Gaps:           25

US-10-008-355-2 (1-712) x AAS55647 (1-2652)

Qy     192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
           |||||||||                       |||       :::|||      |||
Db      76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 135

Qy     204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
           |||:::      ||||||                                       |||
Db     136 CCAAACGTTACAGGTAAACTT---------------------------------CACCTT 162

Qy     224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
           |||                                                   ::::::
Db     163 GGT---------------------------------------------------CACGCT 171

Qy     244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
              |||            :::   :::                  |||||||||:::
Db     172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAAACGCATGCAAGGTTTTGAT 228

Qy     264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                             |||:::      ||||||               :::   |||
Db     229 -----------------ACCCTTTGGCTTCCTGGGATGGACCACGCAGGATTGCCACT 270

Qy     283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                 ||||||:::|||:::         |||         |||   ::::::
Db     271 CAGGCTAAGGTTGAGGAGCGCTTGCGTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT 330

Qy     303 ------------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                       |||      ||||||:::                  :::   ||||||
Db     331 GAGTCTTTCTTGACGAAAGTCTGGGAATGGAAAGAC------------GAATATGCCACT 378

Qy     319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
              ||||||          |||         |||   :::|||   :::
Db     379 ACCATCAAGGAACAATGGGGCAAGATGGGGCTCTCTGTAGACTATTCTCGTGAGCGTTTC 438
```

```
Qy     338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
              ::::::  |||||||:::                      :::|||     :::
Db     439 ACTCTTGACGAAGGTTTGTCA------------------------AAAGCTGTTCGTAAG 474

Qy     358 AlaPheAlaAspTrpIleArgLys------------AsnGlyLysSerAlaValTyrGly 373
            |||   |||      :::|||              |||:::        :::
Db     475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 534

Qy     374 Asp---------------ValLeuSerSerLeuGluLysAlaTyrLys------------ 384
            |||              ||||||   :::|||    :::|||
Db     535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTGGAAGGT 594

Qy     385 ------------------------------GluGlyAlaLysAlaAsnArgGluMetThr 394
                                         :::|||::::::|||         |||
Db     595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGCCCTTGAAGTTGCTACA 654

Qy     395 TyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                     ||||||:::||||||
Db     655 ACTCGTCCTGAGACTATGTTTGGGGACGTT------------------------------ 684

Qy     415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
            |||:::|||   |||||||:::                          |||   :::|||
Db     685 GCGATTGCGGTTAATCCAGAA--------------------------GACCCGCGCTAC 717

Qy     435 LysAsp------------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
            ||||||                  ||||||   :::        |||::::::|||
Db     718 AAGGACTTGATTGGTAAAAACGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy     449 MetLeuAsp---------------------------------------IleValArgArg 455
            :::   |||                                       |||
Db     775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy     456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
                  |||   |||   |||               |||:::
Db     835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACATCATG 894

Qy     476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
            |||            :::  ||||||:::  |||    :::              |||
Db     895 AACGACGACGGAACTATGAACGAGCTTGTCTTTGAATTTTCAGGCATG---------GAT 945

Qy     496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
            :::|||   |||               |||      ||||||      |||
Db     946 CGTTTTGAAGCTCGT---------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy     516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla------------ 531
                     |||::::::   |||   |||   :::      :::
Db     991 GGTGCCCTTGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAACGTACAGGT 1050

Qy     532 ---------------------------------------------------IleGlnAla 534
                                                              :::
Db    1051 GTCGTAGTTGAGCCACGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy     535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
            :::|||:::||||||         :::|||
Db    1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG--------------------------- 1143

Qy     555 ArgGluMetTyrProGlyArgAlaLeuProSer--------------------------- 565
            |||   ||||||   |||      :::
Db    1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203

Qy     566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
            |||         :::   ::::::   :::|||                  |||
Db    1204 GACTGGGTTATCTCTCGTCAGCTCTGGTGGGGTCAC------------------CAAATC 1245

Qy     586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
              ||||||||||||         |||:::     :::   ::::::   |||:::
Db    1246 CCTGCTTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy     606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
              |||   :::        |||::::::||||||
Db    1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335
```

RESULT 6

```
AAX57027
ID   AAX57027 standard; DNA; 2652 BP.
XX
AC   AAX57027;
XX
DT   19-JUL-1999  (first entry)
XX
DE   S. pneumoniae valS coding region DNA.
XX
KW   valS; treatment; antagonist; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   JP11113577-A.
XX
PD   27-APR-1999.
XX
PF   17-OCT-1997;   97JP-0321886.
XX
PR   17-OCT-1997;   97JP-0321886.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
DR   WPI; 1999-320830/27.
DR   P-PSDB; AAY08339.
XX
PT   New valS polypeptide and polynucleotide - useful for treatment of
PT   conditions associated with abnormal valS protein levels
XX
PS   Claim 4; Page 24; 27pp; Japanese.
XX
CC   This invention describes the isolation of a novel Streptococcus
CC   pneumoniae valS protein. The new polypeptide and antagonist are
CC   useful for the treatment of an individual with abnormal valS
CC   protein levels, by administering valS to individuals requiring
CC   valS polypeptide, and administering the antagonist to individuals
CC   requiring inhibition of valS polypeptide.
XX
SQ   Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;

Alignment Scores:
Pred. No.:              0.0223          Length:         2652
Score:                  130.00          Matches:        106
Percent Similarity:     32.81%          Conservative:   62
Best Local Similarity:  20.70%          Mismatches:     166
Query Match:            3.50%           Indels:         178
DB:                     20              Gaps:           25

US-10-008-355-2 (1-712) x AAX57027 (1-2652)

Qy     192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
            |||||||||                       |||      :::|||      |||
Db      76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 135

Qy     204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
            |||:::      ||||||                                     |||
Db     136 CCAAACGTTACAGGTAAACTT---------------------------------CACCTT 162

Qy     224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
            |||                                                   :::::
Db     163 GGT---------------------------------------------------CACGCT 171

Qy     244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
            |||            :::   :::                  |||||||||:::
Db     172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAAACGCATGCAAGGTTTTGAT 228

Qy     264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                             |||:::      ||||||            :::   |||
Db     229 -----------------ACCCTTTGGCTTCCTGGGATGGACCACGCAGGGATTGCCACT 270

Qy     283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                     ||||||:::|||:::      |||         |||   ::::::
Db     271 CAGGCTAAGGTAGAGGAGCGCTTGCGTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT 330
```

```
Qy     303 -----------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                      |||        ||||||:::           :::   ||||||
Db     331 GAGTCTTTCTTGACGAAAGTCTGGGAATGGAAAGAC-----------GAATATGCCACT 378

Qy     319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
              ||||||         |||        |||   :::|||   :::
Db     379 ACTATCAAGGAACAATGGGGCAAGATGGGGCTCTCTGTAGACTATTCTCGTGAGCGTTTC 438

Qy     338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
              ::::::  ||||||:::                        :::|||      :::
Db     439 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 474

Qy     358 AlaPheAlaAspTrpIleArgLys------------AsnGlyLysSerAlaValTyrGly 373
              |||   |||       :::|||               |||:::       :::
Db     475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 534

Qy     374 Asp---------------ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys 388
           |||                ||||||    :::|||      :::|||:::
Db     535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTGGAAGGT 594

Qy     389 AlaAsnArgGluMetThrTyrLeu------------------------------------ 396
           |||         |||   |||:::
Db     595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGTCCTTGAAGTTGCTACA 654

Qy     397 ------SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                 ||||||:::||||||
Db     655 ACTCGTCCTGAGACCATGTTTGGGGACGTT------------------------------ 684

Qy     415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
           |||:::|||   ||||||:::                              |||   :::|||
Db     685 GCGGTTGCGGTCAACCCAGAA--------------------------GACCCGCGCTAC 717

Qy     435 LysAsp-----------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
           ||||||                   ||||||   :::      |||:::::::|||
Db     718 AAGGACTTGATTGGTAAAAATGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy     449 MetLeuAsp---------------------------------------IleValArgArg 455
           :::   |||                                          |||
Db     775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy     456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
                 |||   |||   |||                 |||:::
Db     835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGTCATG 894

Qy     476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
              |||            |||   |||:::   |||   :::            |||
Db     895 AACGACGACGGAACCATGAATGACTTGGCCTTTGAATTTTCAGGCATG---------GAC 945

Qy     496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
           :::|||   |||                |||         ||||||      |||
Db     946 CGTTTTGAAGCTCGT---------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy     516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla------------ 531
                 |||::::::   |||   |||   :::         :::
Db     991 GGTGCCCTCGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAGCGTACAGGT 1050

Qy     532 ---------------------------------------------------IleGlnAla 534
                                                                 :::
Db    1051 GTTGTGGTTGAACCTCGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy     535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
           :::|||:::||||||               :::|||
Db    1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG--------------------------- 1143

Qy     555 ArgGluMetTyrProGlyArgAlaLeuProSer--------------------------- 565
              |||   ||||||   |||                 :::
Db    1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203

Qy     566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
           |||         :::   ::::::   :::|||                      |||
Db    1204 GACTGGGTTATCTCTCGTCAGCTCTGGTGGGGTCAC------------------CAAATC 1245

Qy     586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
               ||||||||||||                |||:::      :::   ::::::   |||:::
Db    1246 CCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy     606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
              |||   :::         |||::::::||||||
Db    1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335
```

```
RESULT 7
AAX32814
ID   AAX32814 standard; DNA; 2652 BP.
XX
AC   AAX32814;
XX
DT   25-JUN-1999  (first entry)
XX
DE   S. pneumoniae valyl tRNA synthetase (ValS) polypeptide encoding DNA.
XX
KW   Valyl tRNA synthetase; valS; otitis media; conjunctivitis; pneumonia;
KW   bacteremia; sinusitis; pleural empyema; endocarditis; meningitis;
KW   bacterial adhesion; bacterial infection; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   EP909818-A1.
XX
PD   21-APR-1999.
XX
PF   17-OCT-1997;   97EP-0308257.
XX
PR   17-OCT-1997;   97EP-0308257.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Brown JR,  Jaworski DD,  Lawlor EJ,  Wang M;
XX
DR   WPI; 1999-231674/20.
DR   P-PSDB; AAY06828.
XX
PT   New Streptococcus pneumoniae valyl tRNA synthetase (valS) gene and
PT   protein, useful for screening for antibacterial compounds, for
PT   prevention and treatment of Streptococci infections e.g. meningitis
PT   and conjunctivitis
XX
PS   Claim 4; Page 6-9; 35pp; English.
XX
CC   This DNA encodes a S. pneumoniae valS polypeptide (EC 6.1.1.9), a member
CC   of the valyl tRNA synthetase family.  The polypeptide can be expressed
CC   recombinantly by transforming host cells with vectors comprising the valS
CC   gene. ValS polypeptides and polynucleotides are useful for diagnosing
CC   diseases due to an infection of an organism and also the stage and type
CC   of the infection. The polypeptides are also used for screening for their
CC   modulators which can be used in the treatment of diseases. Diseases
CC   diagnosed, prevented or treated include otitis media, conjunctivitis,
CC   pneumonia, bacteremia, sinusitis, pleural empyema, endocarditis and
CC   especially meningitis. ValS polypeptides, polynucleotides and their
CC   (ant)agonists can prevent adhesion of bacteria to matrix proteins, and
CC   are useful for use on wounds and body implants to prevent bacterial
CC   infection.
XX
SQ   Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;

Alignment Scores:
Pred. No.:              0.0223        Length:        2652
Score:                  130.00        Matches:       105
Percent Similarity:     32.81%        Conservative:  62
Best Local Similarity:  20.70%        Mismatches:    166
Query Match:            3.50%         Indels:        173
DB:                     20            Gaps:          25

US-10-008-355-2 (1-712) x AAX32814 (1-2652)

Qy     192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
```

```
            ||||||||                               |||          :::|||     |||
Db      76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 135

Qy     204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
           |||:::     ||||||                                        |||
Db     136 CCAAACGTTACAGGTAAACTT--------------------------------CACCTT 162

Qy     224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
           |||                                                      ::::::
Db     163 GGT---------------------------------------------------CACGCT 171

Qy     244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
              |||          :::   :::                     |||||||||:::
Db     172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAAACGCATGCAAGGTTTTGAT 228

Qy     264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                             |||:::    |||||               :::   |||
Db     229 -----------------ACCCTTTGGCTTCCTGGGATGGACCACGCAGGGATTGCCACT 270

Qy     283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                    ||||||:::|||:::        |||          |||   ::::::
Db     271 CAGGCTAAGGTAGAGGAGCGCTTGCGTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT 330

Qy     303 -----------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                          |||      ||||||:::                :::   ||||||
Db     331 GAGTCTTTCTTGACGAAAGTCTGGGAATGGAAAGAC------------GAATATGCCACT 378

Qy     319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
             ||||||          |||          |||    :::|||   :::
Db     379 ACTATCAAGGAACAATGGGGCAAGATGGGGCTCTCTGTAGACTATTCTCGTGAGCGTTTC 438

Qy     338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
              ::::::   |||||||:::                        :::|||     :::
Db     439 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 474

Qy     358 AlaPheAlaAspTrpIleArgLys-----------AsnGlyLysSerAlaValTyrGly 373
              |||   |||      :::|||                |||:::       :::
Db     475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 534

Qy     374 Asp---------------ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys 388
           |||                     ||||||    :::|||        :::||||:::
Db     535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTGGAAGGT 594

Qy     389 AlaAsnArgGluMetThrTyrLeu----------------------------------- 396
           |||         |||   |||:::
Db     595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGTCCTTGAAGTTGCTACA 654

Qy     397 ------SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                    ||||||:::||||||
Db     655 ACTCGTCCTGAGACCATGTTTGGGGACGTT----------------------------- 684

Qy     415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
           |||:::|||   ||||||:::                                 |||   :::|||
Db     685 GCGGTTGCGGTCAACCCAGAA---------------------------GACCCGCGCTAC 717

Qy     435 LysAsp------------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
           ||||||                  ||||||   :::         |||:::::::|||
Db     718 AAGGACTTGATTGGTAAAAATGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy     449 MetLeuAsp-----------------------------------------IleValArgArg 455
           :::   |||                                         |||
Db     775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy     456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
                 |||   |||   |||               |||:::
Db     835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGTCATG 894

Qy     476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
              |||             |||      |||:::   |||   :::         |||
Db     895 AACGACGACGGAACCATGAATGACTTGGCCTTTGAATTTTCAGGCATG---------GAC 945

Qy     496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
           :::|||   |||                  |||        ||||||        |||

Db     946 CGTTTTGAAGCTCGT--------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy     516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla------------ 531
                    |||:::::::   |||   |||     :::          :::
Db     991 GGTGCCCTCGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAGCGTACAGGT 1050

Qy     532 ---------------------------------------------------IleGlnAla 534
                                                              :::
Db    1051 GTTGTGGTTGAACCTCGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy     535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
           :::|||:::||||||               :::|||
Db    1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG--------------------------- 1143

Qy     555 ArgGluMetTyrProGlyArgAlaLeuProSer--------------------------- 565
              |||   ||||||   |||             :::
Db    1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203

Qy     566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
              |||         :::   ::::::   :::|||                  |||
Db    1204 GACTGGGTTATCTCTCGTCAGCTCTGGTGGGGTCAC------------------CAAATC 1245

Qy     586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
              ||||||||||||               |||:::      :::   ::::::   |||:::
Db    1246 CCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy     606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
              |||   :::          |||:::::::||||||
Db    1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335
```

```
RESULT 8
AAA14369
ID   AAA14369 standard; DNA; 2652 BP.
XX
AC   AAA14369;
XX
DT   15-AUG-2000  (first entry)
XX
DE   DNA encoding Streptococcus pneumoniae valyl tRNA synthetase (valS).
XX
KW   Valyl tRNA synthetase; valS; inhibitor; stringent response;
KW   drug screening; antibacterial; antibiotic; genetic immunisation;
KW   antibody; bacterial infection; meningitis; ds.
XX
OS   Streptococcus pneumoniae strain 0100993.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1652
FT                   /*tag=  a
FT                   /product= "Streptococcus pneumoniae valyl tRNA synthetase
FT                   (valS)"
XX
PN   US6051413-A.
XX
PD   18-APR-2000.
XX
PF   24-SEP-1998;   98US-0159539.
XX
PR   18-APR-1996;   96GB-0007791.
PR   17-OCT-1997;   97US-0953492.
PR   18-APR-1997;   97US-0844064.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Brown JR,  Lawlor EJ,  Wang M,  Jaworski DD;
XX
DR   WPI; 2000-338311/29.
DR   P-PSDB; AAY90514.
XX
PT   Novel valS polypeptides of valyl tRNA synthetase family useful for
PT   treating otitis media, conjunctivitis, pneumonia and bacteremia
PT   comprises a specified amino acid sequence -
XX
```

```
PS   Claim 13; Columns 7-10; 21pp; English.
XX
CC   This sequence represents DNA encoding Streptococcus pneumoniae valyl tRNA
CC   synthetase (valS, NCIMB Deposit No. 40794). Streptococci are the cause of
CC   several types of human diseases, including otitis media, conjunctivitis,
CC   pneumonia, bacteraemia, sinusitis, pleural empyema, endocarditis and
CC   especially meningitis. The frequency of S. pneumoniae infections has
CC   risen dramatically over the past 20 years, probably due to the emergence
CC   of multiply antibiotic resistant strains and an increasing population of
CC   immunocompromised people. ValS represents a target for new antibacterial
CC   agents. Inhibition of tRNA synthetases such as valS leads to a reduction
CC   in the levels of charged tRNA, which triggers a cascade of responses
CC   (known as the stringent response) resulting in a state or dormancy in the
CC   bacterium. ValS, its variants and fragments, anti-valS antibodies, valS
CC   inhibitors and nucleotides encoding valS may be used in the diagnosis,
CC   prevention and treatment of bacterial infections such as meningitis.
CC   ValS can be used to screen compounds for inhibitory activity. ValS
CC   may also be useful as an antigen for vaccination of a host to produce
CC   specific antibodies which protect against bacterial invasion into
CC   damaged tissues. Such antibodies could, for example, prevent the
CC   adherence of bacteria to wounds. Nucleotides encoding valS may be used as
CC   diagnostic reagents and therapeutic or prophylactic agents, particularly
CC   for genetic immunisation.
XX
SQ   Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;

Alignment Scores:
Pred. No.:              0.0223        Length:        2652
Score:                  130.00        Matches:       106
Percent Similarity:     32.81%        Conservative:  62
Best Local Similarity:  20.70%        Mismatches:    166
Query Match:            3.50%         Indels:        178
DB:                     21            Gaps:          25

US-10-008-355-2 (1-712) x AAA14369 (1-2652)

Qy     192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
           |||||||||                       |||      :::|||      |||
Db      76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 135

Qy     204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
           |||:::      ||||||                                    |||
Db     136 CCAAACGTTACAGGTAAACTT--------------------------------CACCTT 162

Qy     224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
           |||                                                  ::::::
Db     163 GGT--------------------------------------------------CACGCT 171

Qy     244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
             |||             :::  :::                  |||||||||:::
Db     172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAAACGCATGCAAGGTTTTGAT 228

Qy     264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                             |||:::      ||||||              :::   |||
Db     229 -----------------ACCCTTTGGCTTCCTGGGATGGACCACGCAGGGATTGCCACT 270

Qy     283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                    ||||||:::|||:::      |||         |||   ::::::
Db     271 CAGGCTAAGGTAGAGGAGCGCTTGCGTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT 330

Qy     303 -----------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                      |||      ||||||:::                 :::   ||||||
Db     331 GAGTCTTTCTTGACGAAAGTCTGGGAATGGAAAGAC-----------GAATATGCCACT 378

Qy     319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
               ||||||            |||         |||   :::|||   :::
Db     379 ACTATCAAGGAACAATGGGGCAAGATGGGGCTCTCTGTAGACTATTCTCGTGAGCGTTTC 438

Qy     338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
              ::::::   ||||||:::                        :::|||   :::
Db     439 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 474

Qy     358 AlaPheAlaAspTrpIleArgLys-----------AsnGlyLysSerAlaValTyrGly 373
                ||| |||       :::|||        |||:::      :::
Db     475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 534

Qy     374 Asp--------------ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys 388
           |||                ||||||   :::|||       :::|||:::
Db     535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTGGAAGGT 594

Qy     389 AlaAsnArgGluMetThrTyrLeu------------------------------------ 396
           |||         |||   |||:::
Db     595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGTCCTTGAAGTTGCTACA 654

Qy     397 ------SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                 ||||||:::||||||
Db     655 ACTCGTCCTGAGACCATGTTTGGGGACGTT------------------------------ 684

Qy     415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
           |||:::|||   ||||||:::                            |||   :::|||
Db     685 GCGGTTGCGGTCAACCCAGAA--------------------------GACCCGCGCTAC 717

Qy     435 LysAsp-----------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
           ||||||                  ||||||   :::       |||::::::|||
Db     718 AAGGACTTGATTGGTAAAAATGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy     449 MetLeuAsp--------------------------------------IleValArgArg 455
           :::   |||                                            |||
Db     775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy     456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
                 |||   |||   |||                |||:::
Db     835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGTCATG 894

Qy     476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
              |||             |||      |||:::   |||    :::         |||
Db     895 AACGACGACGGAACCATGAATGACTTGGCCTTTGAATTTTCAGGCATG---------GAC 945

Qy     496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
           :::|||   |||                   |||         |||||   |||
Db     946 CGTTTTGAAGCTCGT---------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy     516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla------------ 531
                 |||::::::   |||   |||      :::         :::
Db     991 GGTGCCCTCGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAGCGTACAGGT 1050

Qy     532 ---------------------------------------------------IleGlnAla 534
                                                               :::
Db    1051 GTTGTGGTTGAACCTCGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy     535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
           :::|||:::||||||                   :::|||
Db    1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG--------------------------- 1143

Qy     555 ArgGluMetTyrProGlyArgAlaLeuProSer--------------------------- 565
              |||   ||||||   |||             :::
Db    1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203

Qy     566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
           |||         :::   ::::::   :::|||                        |||
Db    1204 GACTGGGTTATCTCTCGTCAGCTCTGGTGGGGTCAC-----------------CAAATC 1245

Qy     586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
              ||||||||||||             |||:::      :::   ::::::   |||:::
Db    1246 CCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy     606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
              |||   :::         |||:::::::||||||
Db    1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335

RESULT 9
ABK52414
ID   ABK52414 standard; DNA; 2652 BP.
XX
AC   ABK52414;
XX
```

```
DT   14-AUG-2002  (first entry)
XX
DE   DNA encoding Streptococcus pneumoniae ValS.
XX
KW   ValS; microbial infection; antibacterial; gene; ds.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2651
FT                   /*tag=  a
FT                   /product= "ValS"
XX
PN   JP2002119292-A.
XX
PD   23-APR-2002.
XX
PF   17-OCT-1997; 2001JP-0216648.
XX
PR   17-OCT-1997;   97JP-0321886.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
DR   WPI; 2002-475332/51.
DR   P-PSDB; AAU97883.
XX
PT   Novel valS polynucleotide useful for producing a valS polypeptide, used
PT   in the treatment of valS-related disorders -
XX
PS   Claim 4; Fig 1-3; 24pp; Japanese.
XX
CC   The invention describes an isolated valS polynucleotide encoding a mature
CC   polypeptide the same as that expressed by a valS gene contained in a
CC   deposited strain of Streptococcus pneumoniae, a complementary sequence or
CC   a polynucleotide containing at least 15 continuous bases of the
CC   polynucleotide. The antibody to the encoded protein is used for the
CC   treatment of microbial infection. This sequence encodes ValS derived from
CC   S. pneumoniae.
XX
SQ   Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;

Alignment Scores:
Pred. No.:              0.0223        Length:        2652
Score:                  130.00        Matches:       106
Percent Similarity:     32.81%        Conservative:  62
Best Local Similarity:  20.70%        Mismatches:    166
Query Match:            3.50%         Indels:        178
DB:                     24            Gaps:          25

US-10-008-355-2 (1-712) x ABK52414 (1-2652)

Qy    192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
          ||||||||                        |||      :::|||      |||
Db     76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 135

Qy    204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
          |||:::   ||||||                                       |||
Db    136 CCAAACGTTACAGGTAAACTT-------------------------------CACCTT 162

Qy    224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
          |||                                                   ::::::
Db    163 GGT---------------------------------------------------CACGCT 171

Qy    244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
             |||             :::   :::                  |||||||||:::
Db    172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAAACGCATGCAAGGTTTTGAT 228

Qy    264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                            |||:::     ||||||           :::    |||
Db    229 -----------------ACCCTTTGGCTTCCTGGGATGGACCACGCAGGGATTGCCACT 270

Qy    283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                 ||||||:::|||:::       |||          |||   ::::::
Db    271 CAGGCTAAGGTAGAGGAGCGCTTGCGTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT 330

Qy    303 -----------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                     |||      ||||||:::               :::   |||||
Db    331 GAGTCTTTCTTGACGAAAGTCTGGGAATGGAAAGAC-----------GAATATGCCACT 378

Qy    319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
             ||||||           |||         |||   :::|||   :::
Db    379 ACTATCAAGGAACAATGGGGCAAGATGGGCTCTCTGTAGACTATTCTCGTGAGCGTTTC 438

Qy    338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
             ::::::   ||||||:::                        :::|||      :::
Db    439 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 474

Qy    358 AlaPheAlaAspTrpIleArgLys-----------AsnGlyLysSerAlaValTyrGly 373
             |||   |||      :::|||          |||:::         :::
Db    475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 534

Qy    374 Asp--------------ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys 388
          |||              ||||||   :::|||      :::|||:::
Db    535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTGGAAGGT 594

Qy    389 AlaAsnArgGluMetThrTyrLeu------------------------------------ 396
          |||         |||   |||:::
Db    595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGTCCTTGAAGTTGCTACA 654

Qy    397 ------SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                   ||||||:::||||||
Db    655 ACTCGTCCTGAGACCATGTTTGGGGACGTT---------------------------- 684

Qy    415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
          |||:::|||   ||||||:::                              |||   :::|||
Db    685 GCGGTTGCGGTCAACCCAGAA------------------------GACCCGCGCTAC 717

Qy    435 LysAsp-----------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
          ||||||                 ||||||   :::      |||:::::::|||
Db    718 AAGGACTTGATTGGTAAAAATGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy    449 MetLeuAsp------------------------------------IleValArgArg 455
          :::   |||                                    |||
Db    775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy    456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
             |||   |||   |||               |||:::
Db    835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGTCATG 894

Qy    476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
             |||         |||      |||:::   |||   :::            |||
Db    895 AACGACGACGGAACCATGAATGACTTGGCCTTTGAATTTTCAGGCATG---------GAC 945

Qy    496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
          :::|||   |||                  |||      ||||||      |||
Db    946 CGTTTTGAAGCTCGT--------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy    516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----------- 531
                |||::::::   |||   |||   :::         :::
Db    991 GGTGCCCTCGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAGCGTACAGGT 1050

Qy    532 --------------------------------------------------IleGlnAla 534
                                                            :::
Db   1051 GTTGTGGTTGAACCTCGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy    535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
          :::|||:::||||||          :::|||
Db   1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG------------------------- 1143

Qy    555 ArgGluMetTyrProGlyArgAlaLeuProSer------------------------- 565
             |||   |||:|||   |||      :::
Db   1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203

Qy    566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
          |||         :::   ::::::   :::|||                     |||
Db   1204 GACTGGGTTATCTCTCGTCAGCTCTGGTGGGGTCAC-----------------CAAATC 1245
```

```
Qy     586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
               |||||||||||             |||:::     :::   ::::::   |||:::
Db    1246 CCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy     606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
               |||   :::        |||::::::|||||
Db    1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335

RESULT 10
AAV52160/c
ID   AAV52160 standard; DNA; 6004 BP.
XX
AC   AAV52160;
XX
DT   23-OCT-1998  (first entry)
XX
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:27.
XX
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9818931-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997;   97WO-US19588.
XX
PR   31-OCT-1996;   96US-0029960.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Dougherty BA,  Fannon M;
PI   Kunsch CA,  Rosen CA;
XX
DR   WPI; 1998-272225/24.
XX
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
XX
PS   Claim 1; Page 305-309; 1409pp; English.
XX
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC   recorded on it, or a representative fragment or a sequence at least 95%
CC   identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC   SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC   Streptococcus pneumoniae. The present invention also describes an
CC   isolated nucleic acid molecule encoding a homologue of any of the
CC   fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC   nucleic acid molecule is produced by a process comprising: (a) screening
CC   a genomic DNA library using as a probe a target sequence defined by any
CC   of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC   library which contain sequences that hybridise to the target sequence and
CC   isolating the nucleic acid molecules from the members; or (b) isolating
CC   mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC   molecules whose nucleotide sequence is homologous to amplification
CC   primers derived from the fragment of the S. pneumoniae genome to prime
CC   the amplification and isolating the amplified sequences. The computer
CC   readable medium can be used in a computer-based system for identifying
CC   fragments of the S. pneumoniae genome of commercial importance, or
CC   expression modulating fragments of the S. pneumoniae genome. Products
CC   from the present invention can be used in diagnosis kits and assays, and
CC   pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ   Sequence 6004 BP; 1733 A; 1296 C; 1115 G; 1860 T; 0 other;

Alignment Scores:
Pred. No.:            0.0831        Length:         6004
Score:                    129.00        Matches:        116
Percent Similarity:       33.62%        Conservative:   77
Best Local Similarity:    20.21%        Mismatches:     194
Query Match:              3.47%         Indels:         187
DB:                       19            Gaps:           26

US-10-008-355-2 (1-712) x AAV52160 (1-6004)

Qy     131 LysIleValLysValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMet 150
           ||||||                        |||         |||   |||   :::::::::
Db    4315 AAAATAAATAATACGAATTTGTCTGCAGAAGAAGCTGCTAAGCAAATTCAAAATAAAATA 4256

Qy     151 GluArgLeu-ArgLysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAs 170
              :::::: ||||||          :::   :::||||||:::    |||   |||
Db    4255 AACAAAATAGAGAAAGGACACACACATGTCTAAAGAACTTTCACCTAAATACAATCCAGC 4196

Qy     170 pGluAsnGlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyrPheLeuIleVa 190
             |||   :::              :::   :::                       :::
Db    4195 CGAGGTTGAGGCTGGTCGTTACCAAAAATGG-----------------------CTTGA 4160

Qy     190 lTyrAspValPhe-----------------------LysAspValArgMetValPheAl 202
             |||||||||                      |||         :::|||
Db    4159 TGCGGATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTGATTCC 4100

Qy     202 aProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHi 222
            ||||||:::      ||||||                                  ||
Db    4099 ACCACCAAACGTAACTGGGAAACTT-------------------------------CA 4073

Qy     222 sThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTy 242
            |   |||                                                 ::
Db    4072 CCTTGGT---------------------------------------------------CA 4064

Qy     242 rSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTy 262
            ::::   |||               :::   :::                |||||||||::
Db    4063 CGCTTGGGATACG---ACTTTGCAGGATATCATCATCCGTCAAAAACGCATGCAAGGCTT 4007

Qy     262 rLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu-- 281
            :                  |||:::   ||||||                   :::
Db    4006 TGAT-----------------ACCCTTTGGCTTCCAGGTATGGACCACGCTGGTATTGC 3965

Qy     282 -ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValAr 301
             |||      ||||||:::|||:::        |||         |||   ::::::
Db    3964 GACTCAGGCTAAGGTTGAGGAGCGCTTACGTGGTGAGGGTATTACGCGTTATGACCTAGG 3905

Qy     301 gGly-----------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAl 317
                             |||         ||||||:::                :::   ||
Db    3904 TCGTGAGTCTTTCTTGACGAAGGTCTGGGAATGGAAAGAC-----------GAATATGC 3857

Qy     317 aThrArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSe 336
            ||||  ||||||              |||         |||   :::|||   :::
Db    3856 CACTACCATCAAGGAACAATGGGGCAAGATGGGGCTTTCTGTAGACTATTCTCGTGAGCG 3797

Qy     336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGl 356
                 ::::::   ||||||:::                              :::|||
Db    3796 TTTCACTCTTGATGAAGGTTTGTCA-----------------------AAAGCTGTTCG 3761

Qy     356 uArgAlaPheAlaAspTrpIleArgLys-----------AsnGlyLysSerAlaValTy 372
             :::   |||   :::        :::|||             |||:::         :::
Db    3760 TAAGGTCTTTGTGAACCTTTACAAGAAAGGCTGGATCTATCGTGGTGAGTTTATCATCAA 3701

Qy     372 rGlyAsp---------------ValLeuSerSerLeuGluLysAlaTyrLys-------- 384
              |||                  ||||||   :::|||       :::|||
Db    3700 CTGGGACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTAGA 3641

Qy     385 ---------------------------------GluGlyAlaLysAlaAsnArgGluMe 393
                                            :::|||::::::|||
Db    3640 AGGTGCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGCCCTTGAAGTTGC 3581

Qy     393 tThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAl 413
             |||         ||||||:::||||||
Db    3580 TACAACTCGTCCTGAGACTATGTTTGGGGACGTT------------------------- 3547

Qy     413 aAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLy 433
```

```
              |||:::|||   ||||||:::                           |||   ::
Db     3546 ----GCGGTTGCGGTTAATCCAGAA-------------------------GACCCGCG 3518

Qy      433 sTyrLysAsp-----------------TyrLeuProSerLeuAspArgLysValLeuPr 447
            :|||||||||                     ||||||    :::     |||:::::||
Db     3517 CTACAAGGACTTGATTGGTAAAAACGTCATCCTTCCA---ATCGCTAATAAACTCATCCC 3461

Qy      447 oAlaMetLeuAsp--------------------------------------IleValAr 454
            |  :::  |||                                        |||
Db     3460 AATCGTTGGAGATGAGCACGCAGATCCTGAGCTTGGTACTGGTGTCGTGAAAATCACACC 3401

Qy      454 gArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPh 474
                       |||   |||   |||              |||:::
Db     3400 TGCCCACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGT 3341

Qy      474 eLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSe 494
                  |||                :::        |||:::   |||   :::
Db     3340 CATGAACGACGACGGAACTATGAATGAGCTTGCCTTTGAATTTTCAGGCATG-------- 3289

Qy      494 rAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGl 514
             |||:::|||   |||                 |||       ||||||        ||
Db     3288 -GATCGTTTTGAAGCTCGT--------------AAGGCAGTCGTTGCTAAGTTGGAAGA 3245

Qy      514 uLysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla-------- 531
            |                |||:::::  |||   |||   :::         :::
Db     3244 AATCGGTGCCCTTGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAACGTAC 3185

Qy      532 ----------------------------------------------------IleGl 533
                                                                 :::
Db     3184 AGGTGTCGTAGTTGAGCCACGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGC 3125

Qy      533 nAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGl 553
                :::|||:::|||||                :::|||
Db     3124 TAAGAACGCCATTGCTAACCAAGACACAGAGGACAAG---------------------- 3088

Qy      553 yLeuArgGluMetTyrProGlyArgAlaLeuProSer---------------------- 565
                  |||   ||||||   |||             :::
Db     3087 ----GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGT 3032

Qy      566 ----AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGl 584
                |||        :::   ::::::   :::|||                       ||
Db     3031 CCACGACTGGGTTATCTCTCGTCAGCTTTGGTGGGGTCAC-----------------CA 2990

Qy      584 nAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspPr 604
                |      |||||||||||         |||:::       :::   ::::::   ||
Db     2989 AATCCCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCC 2936

Qy      604 oLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
            |:::   |||   :::         |||:::::||||||
Db     2935 AGAAGGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 2896
```

RESULT 11

```
AAS55949
ID   AAS55949 standard; DNA; 2652 BP.
XX
AC   AAS55949;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Streptococcus pneumoniae DNA for cellular proliferation protein #520.
XX
KW   Antisense; ds; prokaryotic cellular proliferation gene;
KW   antibiotic; antibacterial; drug design.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   P-PSDB; AAU38090.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 27; Seq ID No 9586; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence encodes an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2652 BP; 736 A; 602 C; 621 G; 693 T; 0 other;

Alignment Scores:
Pred. No.:              0.0433        Length:         2652
Score:                  127.00        Matches:        105
Percent Similarity:     33.01%        Conservative:   64
Best Local Similarity:  20.51%        Mismatches:     165
Query Match:            3.41%         Indels:         178
DB:                     23            Gaps:           25

US-10-008-355-2 (1-712) x AAS55949 (1-2652)

Qy      192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
            |||||||||                         |||         :::|||     |||
Db       76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTGATTCCACCA 135

Qy      204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
            |||:::      ||||||                                       |||
Db      136 CCAAACGTAACTGGGAAACTT-------------------------------CACCTT 162

Qy      224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
            |||                                                   ::::::
Db      163 GGT---------------------------------------------------CACGCT 171

Qy      244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
               |||            :::   :::                     |||||||||:::
Db      172 TGGGATACG---ACTTTGCAGGATATCATCATCCGTCAAAAACGCATGCAGGGCTTTGAT 228

Qy      264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                              |||:::   ||||||                   :::   |||
Db      229 -----------------ACCCTTTGGCTTCCAGGTATGGACCACGCTGGTATTGCGACT 270

Qy      283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
```

```
                   ||||||:::|||:::         |||          |||    ::::::
Db      271 CAGGCTAAGGTTGAGGAGCGCTTACGTGGTGAGGGTATTACGCGTTATGACCTAGGTCGT 330

Qy      303 -----------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                       |||        ||||||:::                 :::   ||||||
Db      331 GAGTCTTTCTTGACGAAGGTCTGGGAATGGAAAGAC-----------GAATATGCCACT 378

Qy      319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
                ||||||                |||         |||   :::|||   :::
Db      379 ACCATCAAGGAACAATGGGGCAAGATGGGGCTTTCTGTAGACTATTCTCGTGAGCGTTTC 438

Qy      338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
              ::::::   ||||||:::                        :::|||      :::
Db      439 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 474

Qy      358 AlaPheAlaAspTrpIleArgLys-----------AsnGlyLysSerAlaValTyrGly 373
              |||   |||      :::|||            |||:::       :::
Db      475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTATCGTGGTGAGTTTATCATCAACTGG 534

Qy      374 Asp---------------ValLeuSerSerLeuGluLysAlaTyrLys----------- 384
            |||               ||||||   :::|||        :::|||
Db      535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTAGAAGGT 594

Qy      385 -----------------------------GluGlyAlaLysAlaAsnArgGluMetThr 394
                                         :::|||:::::::|||           |||
Db      595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGCCCTTGAAGTTGCTACA 654

Qy      395 TyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                       ||||||:::||||||
Db      655 ACTCGTCCTGAGACTATGTTTGGGGACGTT----------------------------- 684

Qy      415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
            |||:::|||   ||||||:::                          |||   :::|||
Db      685 GCGGTTGCGGTTAATCCAGAA-------------------------GACCCGCGCTAC 717

Qy      435 LysAsp-----------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
            ||||||                  ||||||   :::      |||:::::|||
Db      718 AAGGACTTGATTGGTAAAAACGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy      449 MetLeuAsp--------------------------------------IleValArgArg 455
            :::   |||                                      |||
Db      775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy      456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
                    |||   |||   |||               |||:::
Db      835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGTCATG 894

Qy      476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
              |||         :::      |||:::   |||   :::         |||
Db      895 AACGACGACGGAACTATGAATGAGCTTGCCTTTGAATTTTCAGGCATG---------GAT 945

Qy      496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
            :::|||   |||                  |||      ||||||      |||
Db      946 CGTTTGAAGCTCGT--------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy      516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----------- 531
                  |||::::::   |||   |||   :::          :::
Db      991 GGTGCCCTTGTCAAAATCGAAAAACGTGTCCATTCAGTCGGTCACTCAGAACGGACAGGT 1050

Qy      532 ------------------------------------------------IleGlnAla 534
                                                            :::
Db     1051 GTCGTAGTTGAGCCACGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy      535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
            :::|||:::||||||             :::|||
Db     1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG------------------------- 1143

Qy      555 ArgGluMetTyrProGlyArgAlaLeuProSer------------------------- 565
              |||   ||||||   |||      :::
Db     1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203

Qy      566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
            |||         :::   ::::::   :::|||                     |||
Db     1204 GACTGGGTTATCTCTCGTCAGCTTTGGTGGGGTCAC-----------------CAAATC 1245

Qy      586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
               |||||||||||                |||:::      :::   ::::::   |||:::
Db     1246 CCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy      606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
               |||   :::         |||::::::||||||
Db     1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335
```

RESULT 12
ABN68148

```
ID   ABN68148 standard; DNA; 4110 BP.
XX
AC   ABN68148;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 4209.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB04789.
XX
PR   27-OCT-2000; 2000GB-0026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J,  Masignani V,  Margarit Ros YI,  Grandi G,  Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP27517.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein -
XX
PS   Claim 7; Page 3577; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence 4110 BP; 1462 A; 620 C; 791 G; 1237 T; 0 other;

Alignment Scores:
Pred. No.:            0.087          Length:        4110
```

```
Score:                    126.50        Matches:        163
Percent Similarity:       33.29%        Conservative:   111
Best Local Similarity:    19.81%        Mismatches:     260
Query Match:              3.40%         Indels:         289
DB:                       24            Gaps:           43

US-10-008-355-2 (1-712) x ABN68148 (1-4110)

Qy     35 AsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeu 54
Db   1831 AATCCAGATAATGAGTCTATTCTGGAAGATATCGTCCAAACTCTAACATTATTTGAAGAC 1890

Qy     55 TyrSerPheAspLysProSerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThr 74
Db   1891 AGAGAAATGATTAAGAAGCGTCTTGAAAACTATAAAGATCTTTTT--------------- 1935

Qy     75 GlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAla 94
Db   1936 ------ACAGAGTCACAACTAAAAAAACTCTATCGTCGTCACTATACTGGCTGGGGACGA 1989

Qy     95 Ile---------------------------------GlnSerGlnSerThrValAspHis 103
Db   1990 TTGTCTGCTAAGTTAATCAATGGTATTCGAGATAAAGAGAGTCAAAAAACAATCTTGGAC 2049

Qy    104 AspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuProIleProGly 123
Db   2050 TATCTTATTGATGATGGTAGATCTAATCGCAACTTTATGCAGTTGATAAATGATGATGGT 2109

Qy    124 LeuSerValLysTyrLeuArgLysIleValLysVal------------------ThrAsp 137
Db   2110 CTATCTTTCAAATCAATTATCAGTAAGGCACAGGCTGGTAGTCATTCAGATAATCTAAAA 2169

Qy    138 LysValGluGlyGlnLeuLysGly------------------------------------ 145
Db   2170 GAAGTTGTAGGTGAGCTTGCAGGTAGCCCTGCTATTAAAAAGGGAATTCTACAAAGTTTG 2229

Qy    146 ---IleThrAspGluMetGluArgLeuArg------LysAlaGlnGluValCysGlnGlu 162
Db   2230 AAAATTGTTGATGAGCTTGTTAAAGTCATGGGATACGAACCTGAACAAATTGTGGTTGAG 2289

Qy    163 LeuAlaLysLysGluAsnAlaAspGluAsnGln--------------------------- 173
Db   2290 ATGGCG---CGTGAGAATCAAACAACAAATCAAGGTCGTCGTAACTCTCGACAACGCTAT 2346

Qy    174 ---------------------------LeuCys------------------IleValGlu 178
Db   2347 AAACTTCTTGATGATGGCGTTAAGAATCTAGCTAGTGACTTGAATGGCAATATTTTGAAA 2406

Qy    179 ProPheTyrSerAsnAsnGlu---------------------------TyrPhe------ 187
Db   2407 GAATATCCTACGGATAATCAAGCGTTGCAAAATGAAAGACTTTTCCTTTACTACTTACAA 2466

Qy    188 ---------------------------------------------LeuIleValTyrAsp 192
Db   2467 AACGGAAGAGATATGTATACAGGGGAAGCTCTAGATATTGACAATTTAAGTCAATATGAT 2526

Qy    193 ValPheLysAspValArgMetValPheAlaProProSerSerValGlyLysPheGlyGly 212
Db   2527 ATTGACCACATTATTCCTCAAGCTTTCATAAAAGATGATTCTATT--------------- 2571

Qy    213 AspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyr 232
Db   2572 ------GATAATCGTGTT------------------------------------TTGGTA 2589

Qy    233 AlaGlyAlaAspAsnArg---------------------ProAlaGluTyrSerLysAsp 245
Db   2590 TCATCTGCTAAAAATCGTGGAAAGTCAGATGATGTTCCTAGCCTTGAAATTGTAAAAGAT 2649

Qy    246 AsnLysProTyr---LysProValTyrPheAlaAlaValSerMetGln------------ 260
Db   2650 TGTAAAGTTTTCTGGAAAAAATTACTTGATGCTAAGTTAATGAGTCAGCGTAAGTATGAT 2709

Qy    261 ------------------------GlyTyrLysAlaAspAspTyrAlaMetThrIleGly 272
Db   2710 AATTTGACTAAGGCAGAGCGCGGAGGCCTAACTTCCGATGATAAGGCAAGATTTATCCAA 2769

Qy    273 PheProGlySerThrAspArgTyrLeuThrSerTrpGly---------ValGluAspArg 289
Db   2770 CGTCAGTTGGTTGAGACACGACAAATTACCAAGCATGTTGCCCGTATCTTGGATGAACGC 2829

Qy    290 IleGluAsnGluAsnAsnPro------------------ArgIleGluValArgGlyIle 303
Db   2830 TTTAATAATGAGCTTGATAGTAAAGGTAGAAGGATCCGCAAAGTTAAAATTGTAACCTTG 2889

Qy    304 LysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLysTyrAla 323
Db   2890 AAGTCAAATTTGGTTTCAAATTTCCGAAAAGAATTTGGATTCTATAAAATTCGTGAAGTT 2949

Qy    324 SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArgGlyLeu 343
Db   2950 AACAATTATCACCATGCACATGATGCCTATCTTAATGCAGTAGTTGCTAAAGCTATTCTA 3009

Qy    344 AlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIle 363
Db   3010 ACCAAATATCCTCAGTTA------------------GAGCCAGAATTT------------ 3039

Qy    364 ArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyr 383
Db   3040 ---------------------GTCTACGGCGACTAT------CCAAAATATAATAGTTAC 3072

Qy    384 LysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGlu--------------- 398
Db   3073 AAAACGCGTAAATCCGCTACAGAAAAGCTATTTTTCTATTCAAATATTATGAACTTCTTT 3132

Qy    399 ------------ThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
Db   3133 AAAACTAAGGTAACTTTAGCGGATGGAACCGTTGTTGTAAAA---------------GAT 3177

Qy    415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
Db   3178 GATATTGAAGTTAATAATGATACGGGT------------GAAATTGTTTGGGATAAAAAG 3225

Qy    435 LysAspTyrLeuProSerLeuAspArgLysValLeu------ProAlaMetLeuAspIle 452
Db   3226 AAACACTTT------GCGACAGTTAGAAAAGTCTTGTCATACCCTCAGAACAATATCGTG 3279

Qy    453 ValArgArgArgIlePro------------------------------------------ 458
Db   3280 AAGAAGACAGAGATTCAGACAGGTGGTTTCTCTAAGGAATCAATCTTGGCGCATGGTAAC 3339

Qy    459 AlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGly------ 476
Db   3340 TCAGATAAGTTG---------------------ATTCCAAGAAAAACGAAGGATATTTAT 3378

Qy    477 ---AspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
Db   3379 TTAGATCCTAAGAAATATGGAGGT------TTTGATAGTCCGATAGTAGCTTACTCTGTT 3432

Qy    496 LysPheHisAlaMetLeuLysSer------------------------------------ 503
Db   3433 TTAGTTGTAGCTGATATCAAAAAGGGTAAAGCACAAAAACTAAAAACAGTTACGGAACTT 3492

Qy    504 ---------------MetAspLysGluLysPheAlaLysAlaIleGluLysAspProAla 518
Db   3493 TTAGGAATTACCATCATGGAGAGGTCCAGATTT------------GAGAAAAATCCATCA 3540

Qy    519 ValGluLeu---SerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMet 537
Db   3541 GCTTTCCTTGAATCAAAAGGCTATTTA------------AATATTAGGGCTGATAAACTA 3588

Qy    538 Ala---------------AsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAla 552
Db   3589 ATTATTTTGCCCAAGTATAGTCTGTTCGAATTAGAAAATGGGCGTCGTCGATTACTTGCT 3648

Qy    553 GlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThr------ 570
```

```
Db   3649 AGTGCTGGTGAATTACAAAAAGGTAATGAGCTAGCCTTACCAACACAATTATGAAGTTC 3708

Qy    571 ------MetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp--------- 585
                :::           |||         ||||||   :::|||::::::
Db   3709 TTATACCTTGCAAGTCGTTATAATGAGTCAAAAGGT---AAACCAGAGGAGATTGAGAAG 3765

Qy    586 ---GlyAlaTrpTyrAsnTyrHisThrThr------------------------------ 594
             :::   |||   |||   :::
Db   3766 AAACAAGAATTTGTAAATCAACATGTCTCTTATTTTGATGACATCCTTCAATTAATTAAT 3825

Qy    595 ------GlyLysGlyVal------------------LeuGluLys--------------- 601
                |||  |||                  |||||||||
Db   3826 GATTTTTCAAAACGAGTTATTCTAGCAGATGCTAATTTAGAGAAAATCAATAAGCTTTAC 3885

Qy    602 GlnAspProLysSer------------AspGluPheAlaValGlnGluAsnIleLeuAsp 617
          ||||||  |||                ||||||   |||        ||||||::::::
Db   3886 CAAGATAATAAGGAAAATATATCAGTAGATGAACTTGCT------AATAATATTATCAAT 3939

Qy    618 LeuPheArgThrLysAsnTyrGly------------------------------------ 625
          ||||||        :::   |||
Db   3940 CTATTTACTTTTACCAGTCTAGGAGCTCCAGCAGCTTTTAAATTTTTGATAAAATAGTT 3999

Qy    626 ---------ArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAsp 642
                   ||||||        ::::::      :::   |||    :::
Db   4000 GATAGAAAACGCTATACATCAACTAAAGAAGTACTTAATTCTACCCTAATTCATCAATCT 4059

Qy    643 IleThrGly 645
          |||||||||
Db   4060 ATTACTGGA 4068
```

RESULT 13
AAZ31948

```
ID   AAZ31948 standard; DNA; 2142 BP.
XX
AC   AAZ31948;
XX
DT   26-JAN-2000  (first entry)
XX
DE   M. catarrhalis strain LES1 tbpB gene.
XX
KW   TbpB gene; Tbp2; transferrin binding protein; diagnosis; otitis media;
KW   genetic immunisation; Moraxella infection; antigen; vaccine; detection;
KW   antitumour antibody production; therapy; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO9952947-A2.
XX
PD   21-OCT-1999.
XX
PF   12-APR-1999;    99WO-CA00307.
XX
PR   14-APR-1998;    98US-0059584.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Myers LE,   Schryvers AB,   Harkness RE,   Loosmore SM,   Du R,   Yang Y;
PI   Klein MH;
XX
DR   WPI; 1999-620376/53.
DR   P-PSDB; AAY43380.
XX
PT   Nucleic acid encoding transferrin binding protein 2 of Moraxella
PT   catarrhalis, useful for diagnostics, immunization and recombinant
PT   protein production  -
XX
PS   Claim 2; Fig 6; 114pp; English.
XX
CC   This sequence encodes the Moraxella catarrhalis strain LES1 transferrin
CC   binding protein (Tbp2) of the invention. This sequence is also referred
CC   to as the TbpB gene. The TbpB gene is used to produce recombinant Tbp2;
CC   for identification or diagnosis of Moraxella, or for cloning related
CC   species, using hybridisation assays; and for genetic immunisation against
CC   Moraxella infections, e.g. otitis media. The Tbp2 proteins are useful as
CC   antigens, either in vaccines (including components of conjugate vaccines
CC   that contain antigens from other bacteria or from tumours, in which case
CC   they elicit production of antitumour antibodies that may be coupled to
CC   chemotherapeutic agents or biologically active agents) or to raise
CC   antibodies (for use as diagnostic reagents and for treating Moraxella
CC   infections), also for detecting Moraxella antibodies.
XX
SQ   Sequence 2142 BP; 753 A; 458 C; 457 G; 474 T; 0 other;

Alignment Scores:
Pred. No.:               0.0705        Length:        2142
Score:                   123.50        Matches:       125
Percent Similarity:      32.17%        Conservative:  77
Best Local Similarity:   19.90%        Mismatches:    273
Query Match:             3.32%         Indels:        153
DB:                      20            Gaps:          25

US-10-008-355-2 (1-712) x AAZ31948 (1-2142)

Qy    119 LeuProIleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLys 138
          :::|||                  :::::||||||:::|||         :::|||      |||
Db    466 ATTCCTTTTGATAAAAATCTTATTGAATACCTTAAAAAATCATCCGAGGTTGTAAGTAAA 525

Qy    139 ValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGlu 158
             |||   |||      ||||||   :::                        |||
Db    526 TTTGAAGCACAAAAAGGCGGTATTGAAAATAACACAAGACTGACACACAAAGATTTATCA 585

Qy    159 ValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGlu 178
                |||:::   ||||||   :::   |||   :::|||   |||
Db    586 TCAGAGCAAAAAGAAGCAAAAGTCAAAGAAGCGTTGGACAATGCTTTA------------ 633

Qy    179 ProPheTyrSerAsnAsnGluTyrPheLeuIleVal---------TyrAspValPheLys 195
                ::::::      :::|||      ::::::            :::|||
Db    634 ACTCAATTTGCCCAAGAAAAATACAAGGAGCTAATTGAGAACGCCCATGATAAAAAATCT 693

Qy    196 AspValArgMetValPheAlaProProSerSerValGlyLysPheGlyGlyAspThrAsp 215
             |||   |||
Db    694 GACGCACGC--------------------------------------------------- 702

Qy    216 AsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAla 235

Db    702 ------------------------------------------------------------ 702

Qy    236 AspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAla 255
             ||||||      ||||||   |||                  |||         |||   |||
Db    703 ---AACCGTGATCTAGAATATGTCAAGTCTGGTTTTAACTATCTTTCTGGATATACCGCC 759

Qy    256 Ala---------------ValSerMetGlnGly------------------TyrLysAla 264
                               :::   :::|||                    ||||||
Db    760 ACCGACCACGACAAAAAAACCAATTATCGTGGCTATTATGGTGCGTTGTATTATAAAGGC 819

Qy    265 AspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrp 284
             :::   |||         :::         |||   ::::::   :::|||         |||
Db    820 AGCGAAACCGCCAAAGAGCTA------CCACAAACAAGTGCAAAATATAAAGGTTATTGG 873

Qy    285 Gly------ValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                    :::      ::::::|||:::                  ::::::   |||
Db    874 GACTTTATGACAGATGCCACACTTGATAACAAATACACG---------GATTTGCCAGGT 924

Qy    303 Ile---LysGlnGlyIleTrpLysGluAlaMetSerAlaAspGln---AlaThrArgIle 320
          |||   :::|||      |||:::      :::|||   |||:::   ||||||   :::
Db    925 ATCGCCAGACAAACCCAGTGGCGTAGTCTTGTTTCTACTGATGAGTATGCAACGCTCTTG 984

Qy    321 ------------------------LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyr 332
                                  |||         |||      |||:::   :::
Db    985 ACAGACAAAAATAACAAGCCCAGTGATTACAATGGTGCATATGGTCATAGCAGTGAATTT 1044

Qy    333 TrpLysAsnSerIleGly---------------------MetAsnArgGlyLeuAlaArg 345
             |||                                  ::::::      |||:::
Db   1045 GATGTTAATTTTGCTGATAAAAAAATTAAAGGCAAACTTATCAGTAATCAGTTATCAGGC 1104

Qy    346 LeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLys 365
```

```
                |||      ::::::|||   :::      |||               |||
Db   1105 ACAGCTGTAACCGCCAAAGAGCGTTATAAAATAGAAGCTGATATCCACGGCAACCGCTTC 1164

Qy    366 AsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGlu 385
             |||   :::                     ::::::   :::||||||      |||
Db   1165 CGTGGCAGTGCC--------------------ACCGCAAGCGATAAAGCA------GAA 1197

Qy    386 GlyAlaLysAlaAsnArgGluMetThrTyr---LeuSerGluThrLeuPheGlyGlyThr 404
             :::|||               |||         :::      |||   ||||||
Db   1198 GACAGCAAAACCCAACACCCCTTTACCAGCGATGCTACAAACAAGCTAGAAGGTGGTTTT 1257

Qy    405 GluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla 424
                   :::      :::   |||           |||:::    :::
Db   1258 TATGGACCAAAAGGCGAGGAGCTGGCAGGTAAATTCTTAACCGATGACAACAAACTCTTT 1317

Qy    425 GlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLys 444
          |||:::      :::      ||||||   :::               ||||||
Db   1318 GGGGTCTTTGGTGCTAAACGAGATAAAGTAGAAAAAACCGAAGCCATCTTAGATGCCTAT 1377

Qy    445 ValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeuProAsp 464
             |||
Db   1378 GCACTT---------------------------------------------------GGG 1386

Qy    465 IlePheLysAsnValIleAsp---LysLysPheLysGlyAspThrLysLysTyrAlaAsp 483
             |||   |||                  |||         |||||||||       |||
Db   1387 ACATTTAATAATACAAATAAAGCAACCACATTCACCCCATTTACCAAAAAACAACTGGAT 1446

Qy    484 ---PheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLys 502
             |||         |||   :::|||      |||         :::   :::
Db   1447 AACTTTGGCAATGCCAAAAAGTTGGTCTTGGGTTCTACCGTCATTAATTTGGTGTCTACC 1506

Qy    503 SerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer 522
                   |||   :::|||   |||         ||||||
Db   1507 GATGCCACCAAAAATGAATTCACCAAAAAATTCACCAAAGAC------------------ 1548

Qy    523 LysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAla 542
          |||         |||      :::|||   :::         |||   |||
Db   1549 AAGCCAACTTCTGCCACAAACAAAGCGGGCGAGACT---TTGATGGTGAATGATGAAGTT 1605

Qy    543 IleGluLys------GlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560
          |||   |||      ||||||                        |||:::      |||
Db   1606 ATCGTCAAAACCTATGGCAAAAACTTTGAATACCTAAAATTTGGTGAGCTTAGTGTCGGT 1665

Qy    561 ArgAla------------LeuProSerAspAlaAsnPheThrMetArgMetSerTyrGly 576
             :::            |||      :::         |||
Db   1666 GATAGCCATAGCGTCTTTTTACAAGGCGAACGCACCGCTACCACAGGCGAGAAAGCCGTA 1725

Qy    577 SerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThr--------- 593
                   |||               |||   |||      |||   |||
Db   1726 CCAACCACAGGCAAAGCCAAATATCTGGGGAACTGGGTAGGATACATCACAGGAGCGGGC 1785

Qy    594 ThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGlu 613
          |||||||||         |||   ||||||                     :::   :::
Db   1786 ACAGGAAAAAGCTTTAATGAGGCCCAAGAT---------------------ATTGCTGAT 1824

Qy    614 AsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeu 633
                :::|||      |||            |||:::            |||:::
Db   1825 TTTGACATTGACTTTGAGAGAAAATCAGTTAAAGGCAAACTGACCACCCAAGGCCGCACA 1884

Qy    634 HisIleAlaPheLeuSerAsnAsnAspIleThrGly------------------------ 645
                   |||            :::|||   |||
Db   1885 GATCCTGTCTTTAACATCAAAGGTGAAATTGCAGGCAATGGCTGGACAGGCAAAGCCAGC 1944

Qy    646 ------------------------------------------GlyAsnSerGlySerProVal 652
                                                    :::::::|||     :::
Db   1945 ACCACCAAAGCGGACGCAGGAGGCTACAAGATAGATTCTAGCAGTACAGGCAAATCCATC 2004

Qy    653 PheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSer 672
                :::|||      :::   |||   :::         |||         |||
Db   2005 GTCATCGAAAATGCCGAAGTTACTGGGGGCTTTTATGGTCCAAATGCAAACGAGATGGGC 2064

Qy    673 GlyAspIleGluPheGluProAsp 680
          |||            :::   |||
Db   2065 GGGTCATTTACACACGATACCGAT 2088
```

RESULT 14
ABQ69595
ID   ABQ69595 standard; DNA; 4716 BP.
XX
AC   ABQ69595;
XX
DT   29-AUG-2002 (first entry)
XX
DE   Listeria innocua DNA sequence #1034.
XX
KW   Antibacterial; Listeria; food contamination; mutational analysis;
KW   infection; ds.
XX
OS   Listeria innocua.
XX
PN   WO200228891-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-FR03061.
XX
PR   04-OCT-2000; 2000FR-0012697.
XX
PA   (INSP) INST PASTEUR.
PA   (CNRS) CNRS CENT NAT RECH SCI.
XX
PI   Kunst F, Glaser P;
XX
DR   WPI; 2002-332479/37.
XX
PT   New genomic sequences from Listeria species, useful for detection,
PT   treatment and prevention of infection, also related polypeptides,
PT   antibodies and modulators -
XX
PS   Claim 7; SEQ ID 2408; 180pp; French.
XX
CC   The present invention relates to nucleic acid sequences
CC   (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC   and primers for identification and/or detection of Listeria (e.g. as
CC   contaminants in foods, or mutational analysis) and for analysis of
CC   gene expression. Proteins encoded by the nucleic acid sequences can be
CC   used to screen for compounds that modulate gene expression, replication
CC   and pathogenicity of Listeria (potential therapeutic agents), also for
CC   treating infections by Listeria, and are useful as immunogens in
CC   anti-Listeria vaccines.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4716 BP; 1750 A; 734 C; 1110 G; 1122 T; 0 other;

Alignment Scores:

Pred. No.:	0.253	Length:	4716
Score:	122.50	Matches:	138
Percent Similarity:	32.84%	Conservative:	104
Best Local Similarity:	18.72%	Mismatches:	308
Query Match:	3.29%	Indels:	187
DB:	24	Gaps:	31

US-10-008-355-2 (1-712) x ABQ69595 (1-4716)

```
Qy     11 GlyAlaAlaLeuLeuLeuGlyAlaSerGlyValAlaLysAlaAsp--------------- 25
          |||      :::|||   |||      |||      |||
Db   1675 GGCGGAGTTGTATTGCCTGGGGAAAAGGAAAAGGGAAAGTTAAACTTTTCAGCAAAGCT 1734

Qy     26 ------------LysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArg 41
                      |||||||||            :::|||
Db   1735 ACTGCAGGACTCAAAGGCATGGGTAAATTCGGTAAGTTG-----------GCTAGCGGT 1782

Qy     42 MetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysProSer 61
          :::      ||||||      |||            |||         :::||| . :::
```

```
Db   1783 GTAGGTGTTTTAGGTGTAGGATTAAGCGCTACTGAGTTGATTGGTATGAACAAAAATAAT 1842

Qy     62 IleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAspGln 81
                                    ||||||||         |||      ::::::
Db   1843 GCAGGTGAAAAAACTGGTGGCTTTGGTGGTTCTTTAGGTGGT------ATGGCTGGTGGT 1896

Qy     82 GlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAla-------------------- 94
                |||   |||            |||   |||
Db   1897 GCTGCAATAGGAACTATGATTGCTCCTGGAGTAGGTACAGCTATTGGTGGGGCAATTGGC 1956

Qy     95 ----------------------------------------IleGlnSerGlnSerThr 100
                                                  :::|||    :::
Db   1957 GCTTTTGCTGGAACTGCATTAGGTAAAGAATTAGGGAAATATGTTCAAAAAGAAGGACCA 2016

Qy    101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThr---MetGlyGluGluLeu 119
                   |||       :::   |||:::         :::   :::   ||||||
Db   2017 AAAATATTAGATAAATTCAAGACAGGTTGGAAAGGCTTAAGCAAAATTGCTGAAGAGCAT 2076

Qy    120 ProIleProGlyLeuSerValLysTyrLeuArgLysIleValLys--------------- 134
          ||||||   |||    ::::::       :::.  |||   :::
Db   2077 CCGATTTTAGGAGCAAATATAAATGTAATAAATAAAACTATCGATGCGGCAAAGAAAGGC 2136

Qy    135 -----------------------------------------------------ValThr 136
                                                               |||
Db   2137 ATTAAAGCAGTTGGAGACACGCATAAAACTGTTTGGAATGCTTCAAAATCAATAGTCGCC 2196

Qy    137 Asp------LysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArg 154
          |||      |||::::::         ||||||::::::    :::
Db   2197 GACCCGCTCAAAATTGATGCGAGCAGTAAAGGTGTTTCTAAAGAC--------------- 2241

Qy    155 LysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeu 174
             |||:::   :::   :::   |||      ::::::         ::::::::::::
Db   2242 TCTGCTAAAGCAATGAACGAGTATTTACGAAATGAGCAGAAAATGCAAGATAGTCGAGTC 2301

Qy    175 --------------------CysIleValGluProPheTyrSerAsnAsnGluTyrPhe 187
                                |||   |||         ||||||
Db   2302 GAAATCATGGTATCTGGTCGTCGTATTACTGAAAAAGAACTCGAAAACAAC--------- 2352

Qy    188 LeuIleValTyrAspValPheLysAspValArgMetValPheAlaProProSerSerVal 207
          :::      ||||||         |||   :::::::::   |||
Db   2353 ATAAAAACATATGATAAAATGTCTGAT---CAGTTAATAGCTGCAACTGAAAAAAAATCA 2409

Qy    208 GlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSer 227
             |||      |||                  |||   :::                 :::
Db   2410 GCAAAAGCAAATAAAGAT---------------TGGGATAAGCTAGTAGCACTTGGGGCT 2454

Qy    228 ValPheArgValTyrAlaGlyAla-----------AspAsnArgProAlaGluTyrSer 243
          |||   :::      |||   |||             ::::|||      |||      :::
Db   2455 GTTTCTAAAGATACAGCTAGCGCAAAGAAACACACTAATAATGAAACTGCATCAATAAAT 2514

Qy    244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
          |||   :::                              |||:::   :::   ||||||
Db   2515 AAAGCGGATATTAAAGCGAATAACGAAGAATTAAAAGCTTTAGAAAGAGAATTCTATAAA 2574

Qy    264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSer 283

Db   2574 ------------------------------------------------------------ 2574

Qy    284 TrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGlyIle 303
             |||   |||:::         |||         |||          :::      |||
Db   2575 ---GGGCAAGAAGAGATTACACGAAATGCTGAAAAT---------AGAATTAATGCTATT 2622

Qy    304 LysGlnGlyIleTrpLysGlu------AlaMetSerAlaAspGlnAlaThrArgIleLys 321
          ||||||         ||||||            ::::::   |||:::      |||:::
Db   2623 AAACAGAAAGCTAGCAAAGAGAAACGGAAATTAACAAAAGATGAAGAAAAAGAAATCCAA 2682

Qy    322 TyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArg 341

Db   2682 ------------------------------------------------------------ 2682

Qy    342 GlyLeuAlaArgLeuAsp------ValIleGlyArgLysArgAlaGluGluArgAlaPhe 359
                   |||::::::            :::      :::|||:::|||   |||::::::
Db   2683 ---------AGAATGGAAACAACAACACTTGAGTTCCGTCGTTCTGAACGACGTTCTTAT 2733

Qy    360 AlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeu 379
                :::||||||
Db   2734 GAAAAAGAAGTTAGAAAG----------------------------------ATTGAA 2757

Qy    380 GluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThr 399
          ||||||      ||||||   |||   |||                           :::
Db   2758 GAAAAGCAAAGGAAAGAAGCTGCTATTGCACTCACTGCTTCAGCTAAAGAACAAAAAATA 2817

Qy    400 LeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThr--- 418
          :::   |||   |||      :::               |||   |||   |||:::
Db   2818 ATACTAGGAAACTTAGAAAACTCGAAAGAAAAAATGTCAGCAAAAGCAGCAGCAAGTGTT 2877

Qy    419 ---------AsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLys 435
                   :::   |||         :::|||      :::   :::||||||
Db   2878 GTTAAAAATTCCGCAAAAGCTAGAGATGCATCCGTAAAAGAAGCTAATAAAGAGTATAAA 2937

Qy    436 AspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArg 455
                            ||||||   |||                     |||
Db   2938 CAGACTAAAAAAATACTTGATGAAAAA-----------------CGATTTGTCACAGGT 2979

Qy    456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
             |||   ::::::      |||      ||||||            ||||||   |||
Db   2980 GAAATTTCAGAAGAAGAATATCAAGATGCGCTGAAAAATGCG------AAGAAAAAGAAA 3033

Qy    476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
                   |||   |||:::   :::   |||      |||
Db   3034 AACGGCGTAGTTAAAGAAGCGGAGAAGATGCACGATAATGTAGTGCGAGAAGCTCAGCAA 3093

Qy    496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
          :::               |||:::   :::|||   |||         :::   :::   |||
Db   3094 CAAGCAAAAGGACATTTAAGAGAAGTAGATTGGGAAACT---GGTGAAACATTAAATAAG 3150

Qy    516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAsp 535
                :::.      |||      :::               :::|||:::
Db   3151 ------TGGGAGCAATTTAAAAAAGGAACAGTTGAAACATTTAACAGTATAAAAGATGCA 3204

Qy    536 AlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArg 555
          |||:::                  :::                     |||:::|||::::::
Db   3205 GCGCTCGGAAAATGGAACGAGTTGTCCTCTGCGACAGTTAATATTTCAGTGGAATGAAA 3264

Qy    556 GluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyr 575
          :::               |||         |||                  :::::
Db   3265 GATACA---GCGTTGAGAGTATTTGATAGTTTTAAGACAAACTTGTATAAAGTCATTAAT 3321

Qy    576 GlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThr------ 593
                |||   |||            :::      :::::::|||         |||
Db   3322 AATGTAATTACCGGAGTCAAT---AAAGTACTATCGTTCTTTAACATTAAGACAATACCT 3378

Qy    594 ---ThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 612
                |||   |||      :::                                    |||
Db   3379 GTGCTTGGGAACGGTAGCATAGGTTCAAGT--------------------------CAA 3411

Qy    613 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGln 632
          ||||||   |||                     |||   |||      :::::::::|||
Db   3412 GAAAACAGTTTATCTGCAAAGGATAAGAAAACGTATCATTCTACATCCCAATCAGGAAAT 3471

Qy    633 LeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly------------GlyAsnSer 648
          |||   :::   :::      :::|||:::   :::|||              |||
Db   3472 CTAGCTATGAACTATACTGGTTCTAATAATGCAAGTGGCCAAATTATGGCTGGTGAGGAA 3531

Qy    649 GlySerProVal---PheAspLysAsn---------GlyArgLeuIleGly 661
          |||      :::      ::::::||||||            |||::::::|||
Db   3532 GGTTTCGAGATTGCATATAACAAGAACTCAGCACAAGCGAGAATTCTAGGG 3582
```

RESULT 15
ABQ67873
ID   ABQ67873 standard; DNA; 4731 BP.
XX
AC   ABQ67873;
XX
DT   29-AUG-2002 (first entry)
XX

```
DE   Listeria innocua DNA sequence #675.
XX
KW   Antibacterial; Listeria; food contamination; mutational analysis;
KW   infection; ds.
XX
OS   Listeria innocua.
XX
PN   WO200228891-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-FR03061.
XX
PR   04-OCT-2000; 2000FR-0012697.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Kunst F,  Glaser P;
XX
DR   WPI; 2002-332479/37.
XX
PT   New genomic sequences from Listeria species, useful for detection,
PT   treatment and prevention of infection, also related polypeptides,
PT   antibodies and modulators  -
XX
PS   Claim 7; SEQ ID 686; 180pp; French.
XX
CC   The present invention relates to nucleic acid sequences
CC   (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC   and primers for identification and/or detection of Listeria (e.g. as
CC   contaminants in foods, or mutational analysis) and for analysis of
CC   gene expression. Proteins encoded by the nucleic acid sequences can be
CC   used to screen for compounds that modulate gene expression, replication
CC   and pathogenicity of Listeria (potential therapeutic agents), also for
CC   treating infections by Listeria, and are useful as immunogens in
CC   anti-Listeria vaccines.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4731 BP; 1752 A; 735 C; 1117 G; 1127 T; 0 other;

Alignment Scores:
Pred. No.:               0.254       Length:         4731
Score:                   122.50      Matches:        138
Percent Similarity:      32.84%      Conservative:   104
Best Local Similarity:   18.72%      Mismatches:     308
Query Match:             3.29%       Indels:         187
DB:                      24          Gaps:           31

US-10-008-355-2 (1-712) x ABQ67873 (1-4731)

Qy     11 GlyAlaAlaLeuLeuLeuGlyAlaSerGlyValAlaLysAlaAsp--------------- 25
             |||       :::|||   |||       |||      |||
Db   1690 GGCGGAGTTGTATTGCCTGGGGAAAAGGAAAAGGGAAAGTTAAACTTTTCAGCAAAGCT 1749

Qy     26 ------------LysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArg 41
                      |||||||||               :::|||
Db   1750 ACTGCAGGACTCAAAGGCATGGGTAAATTCGGTAAGTTG-----------GCTAGCGGT 1797

Qy     42 MetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysProSer 61
          :::      ||||||      |||            |||         :::|||   :::
Db   1798 GTAGGTGTTTTAGGTGTAGGATTAAGCGCTACTGAGTTGATTGGTATGAACAAAAATAAT 1857

Qy     62 IleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAspGln 81
                               |||||||||         |||        ::::::
Db   1858 GCAGGTGAAAAAACTGGTGGCTTTGGTGGTTCTTTAGGTGGT------ATGGCTGGTGGT 1911

Qy     82 GlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAla--------------------- 94
                |||   |||               |||   |||
Db   1912 GCTGCAATAGGAACTATGATTGCTCCTGGAGTAGGTACAGCTATTGGTGGGGCAATTGGC 1971

Qy     95 -----------------------------------------IleGlnSerGlnSerThr 100
                                                  :::|||   :::
Db   1972 GCTTTTGCTGGAACTGCATTAGGTAAAGAATTAGGGAAATATGTTCAAAAAGAAGGACCA 2031

Qy    101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThr---MetGlyGluGluLeu 119
             |||         :::   |||:::        :::   :::   ||||||
Db   2032 AAAATATTAGATAAATTCAAGACAGGTTGGAAAGGCTTAAGCAAAATTGCTGAAGAGCAT 2091

Qy    120 ProIleProGlyLeuSerValLysTyrLeuArgLysIleValLys--------------- 134
          ||||||   |||      ::::::         :::   |||   :::
Db   2092 CCGATTTTAGGAGCAAATATAAATGTAATAAATAAAACTATCGATGCGGCAAAGAAAGGC 2151

Qy    135 ------------------------------------------------------ValThr 136
                                                                   |||
Db   2152 ATTAAAGCAGTTGGAGACACGCATAAAACTGTTTGGAATGCTTCAAAATCAATAGTCGCC 2211

Qy    137 Asp------LysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArg 154
          |||         |||::::::            ||||||::::::   :::
Db   2212 GACCCGCTCAAAATTGATGCGAGCAGTAAAGGTGTTTCTAAAGAC--------------- 2256

Qy    155 LysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeu 174
             |||:::   :::      :::   |||               ::::::::::::
Db   2257 TCTGCTAAAGCAATGAACGAGTATTTACGAAATGAGCAGAAAATGCAAGATAGTCGAGTC 2316

Qy    175 ---------------------CysIleValGluProPheTyrSerAsnAsnGluTyrPhe 187
                                  |||   |||         ||||||
Db   2317 GAAATCATGGTATCTGGTCGTCGTATTACTGAAAAAGAACTCGAAAACAAC--------- 2367

Qy    188 LeuIleValTyrAspValPheLysAspValArgMetValPheAlaProProSerSerVal 207
          :::      ||||||            |||   :::::::::   |||
Db   2368 ATAAAAACATATGATAAAATGTCTGAT---CAGTTAATAGCTGCAACTGAAAAAAAATCA 2424

Qy    208 GlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSer 227
             |||         |||                  |||   :::               :::
Db   2425 GCAAAAGCAAATAAAGAT---------------TGGGATAAGCTAGTAGCACTTGGGGCT 2469

Qy    228 ValPheArgValTyrAlaGlyAla------------AspAsnArgProAlaGluTyrSer 243
          |||   :::         |||   |||            :::|||      |||      :::
Db   2470 GTTTCTAAAGATACAGCTAGCGCAAAGAAACACACTAATAATGAAACTGCATCAATAAAT 2529

Qy    244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
          |||   :::                           |||:::      :::   ||||||
Db   2530 AAAGCGGATATTAAAGCGAATAACGAAGAATTAAAAGCTTTAGAAAGAGAATTCTATAAA 2589

Qy    264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSer 283

Db   2589 ------------------------------------------------------------ 2589

Qy    284 TrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGlyIle 303
             |||   |||:::            |||      |||         :::      |||
Db   2590 ---GGCAAGAAGAGATTACACGAAATGCTGAAAAT---------AGAATTAATGCTATT 2637

Qy    304 LysGlnGlyIleTrpLysGlu------AlaMetSerAlaAspGlnAlaThrArgIleLys 321
          ||||||         ||||||            ::::::   |||:::         |||:::
Db   2638 AAACAGAAAGCTAGCAAAGAGAAACGGAAATTAACAAAAGATGAAGAAAAAGAAATCCAA 2697

Qy    322 TyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArg 341

Db   2697 ------------------------------------------------------------ 2697

Qy    342 GlyLeuAlaArgLeuAsp------ValIleGlyArgLysArgAlaGluGluArgAlaPhe 359
                   |||::::::         :::      :::|||:::|||   |||::::::
Db   2698 ---------AGAATGGAAACAACAACACTTGAGTTCCGTCGTTCTGAACGACGTTCTTAT 2748

Qy    360 AlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeu 379
                 :::||||||
Db   2749 GAAAAAGAAGTTAGAAAG------------------------------------ATTGAA 2772

Qy    380 GluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThr 399
          ||||||         ||||||      |||   |||                     :::
Db   2773 GAAAAGCAAAGGAAAGAAGCTGCTATTGCACTCACTGCTTCAGCTAAAGAACAAAAAATA 2832

Qy    400 LeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThr--- 418
```

```
           :::   |||      |||      :::            |||   |||   |||:::
Db    2833 ATACTAGGAAACTTAGAAAACTCGAAAGAAAAAATGTCAGCAAAAGCAGCAGCAAGTGTT 2892

Qy     419 ---------AsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLys 435
                    :::      |||            :::|||      :::   :::||||||
Db    2893 GTTAAAAATTCCGCAAAAGCTAGAGATGCATCCGTAAAAGAAGCTAATAAAGAGTATAAA 2952

Qy     436 AspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArg 455
                          ||||||   |||                        |||
Db    2953 CAGACTAAAAAAATACTTGATGAAAAA-----------------CGATTTGTCACAGGT 2994

Qy     456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
              |||      :::::       |||      ||||||         ||||||   |||
Db    2995 GAAATTTCAGAAGAAGAATATCAAGATGCGCTGAAAAATGCG------AAGAAAAAGAAA 3048

Qy     476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
                       |||   |||:::   :::   |||      |||
Db    3049 AACGGCGTAGTTAAAGAAGCGGAGAAGATGCACGATAATGTAGTGCGAGAAGCTCAGCAA 3108

Qy     496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
           :::            |||:::   :::|||   |||         :::   :::   |||
Db    3109 CAAGCAAAAGGACATTTAAGAGAAGTAGATTGGGAAACT---GGTGAAACATTAAATAAG 3165

Qy     516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAsp 535
                       :::      |||      :::            :::|||:::
Db    3166 ------TGGGAGCAATTTAAAAAAGGAACAGTTGAAACATTTAACAGTATAAAAGATGCA 3219

Qy     536 AlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArg 555
           |||:::               :::                     |||:::|||::::::
Db    3220 GCGCTCGGAAAATGGAACGAGTTGTCCTCTGCGACAGTTAATATTTCAGTGGAATGAAA 3279

Qy     556 GluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyr 575
           :::            |||         |||                  ::::::
Db    3280 GATACA---GCGTTGAGAGTATTTGATAGTTTTAAGACAAACTTGTATAAAGTCATTAAT 3336

Qy     576 GlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThr------ 593
                 |||   |||         :::      :::::::::|||      |||
Db    3337 AATGTAATTACCGGAGTCAAT---AAAGTACTATCGTTCTTTAACATTAAGACAATACCT 3393

Qy     594 ---ThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 612
                 |||   |||   :::                                    |||
Db    3394 GTGCTTGGGAACGGTAGCATAGGTTCAAGT---------------------------CAA 3426

Qy     613 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGln 632
           ||||||   |||               |||   |||         :::::::::|||
Db    3427 GAAAACAGTTTATCTGCAAAGGATAAGAAAACGTATCATTCTACATCCCAATCAGGAAAT 3486

Qy     633 LeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly------------GlyAsnSer 648
           |||   :::   :::      :::|||:::   :::|||            |||
Db    3487 CTAGCTATGAACTATACTGGTTCTAATAATGCAAGTGGCCAAATTATGGCTGGTGAGGAA 3546

Qy     649 GlySerProVal---PheAspLysAsn---------GlyArgLeuIleGly 661
           |||      :::   ::::::||||||            |||::::::|||
Db    3547 GGTTTCGAGATTGCATATAACAAGAACTCAGCACAAGCGAGAATTCTAGGG 3597
```

Search completed: May 23, 2003, 10:05:47
Job time : 379.182 secs

P
D


```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         May 23, 2003, 09:42:03 ; Search time 60.7132 Seconds
                                          (without alignments)
                                          3596.482 Million cell updates/sec

Title:          US-10-008-355-2
Perfect score:  3719
Sequence:       1 MQMKLKSILLGAALLLGASG..........LFMIDKWGQCPRLIQELKLI 712

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters:      882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110403_2582/app_query.fasta_1.1230
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_@CGN_1_1_47_@runat_16052003_110403_2582 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2995	80.5	1974	4	US-09-221-017B-726	Sequence 726, App
	2	743.5	20.0	1317	4	US-09-221-017B-382	Sequence 382, App
	3	441.5	11.9	2384	4	US-09-221-017B-1045	Sequence 1045, Ap
	4	265.5	7.1	561	4	US-09-221-017B-12	Sequence 12, Appl
	5	130	3.5	2652	2	US-08-953-492-1	Sequence 1, Appli
c	6	129	3.5	6004	4	US-08-961-527-27	Sequence 27, Appl
	7	123.5	3.3	2142	4	US-09-059-584-52	Sequence 52, Appl
	8	115	3.1	4853	2	US-08-793-824-1	Sequence 1, Appli
c	9	111	3.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	10	110.5	3.0	12665	4	US-08-961-527-134	Sequence 134, App
	11	110	3.0	2319	4	US-09-134-078-14	Sequence 14, Appl
	12	110	3.0	2667	4	US-09-134-001C-195	Sequence 195, App
	13	110	3.0	3279	4	US-08-446-137B-1	Sequence 1, Appli
c	14	109	2.9	15363	4	US-08-961-527-139	Sequence 139, App
c	15	109	2.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	16	107.5	2.9	4851	4	US-09-221-017B-999	Sequence 999, App
	17	107	2.9	38155	4	US-09-453-702B-79	Sequence 79, Appl
c	18	106.5	2.9	15213	4	US-08-961-527-26	Sequence 26, Appl
	19	106	2.9	2472	3	US-08-335-844A-7	Sequence 7, Appli
	20	106	2.9	3358	3	US-08-335-844A-20	Sequence 20, Appl
	21	106	2.9	3698	4	US-09-327-536-1	Sequence 1, Appli
	22	105.5	2.8	1994	3	US-08-600-982-22	Sequence 22, Appl
	23	105.5	2.8	1994	5	PCT-US94-10261A-22	Sequence 22, Appl
	24	104	2.8	3731	4	US-08-811-583-1	Sequence 1, Appli
	25	102.5	2.8	2571	4	US-09-134-001C-2601	Sequence 2601, Ap
	26	102.5	2.8	6142	4	US-09-514-302-1	Sequence 1, Appli
	27	102	2.7	1242	4	US-09-134-001C-2274	Sequence 2274, Ap
	28	102	2.7	1415	4	US-09-345-214-11	Sequence 11, Appl
	29	102	2.7	2008	4	US-09-345-214-12	Sequence 12, Appl
	30	102	2.7	2491	4	US-09-345-214-5	Sequence 5, Appli
	31	102	2.7	2582	2	US-08-816-105A-2	Sequence 2, Appli
	32	102	2.7	4310	3	US-09-008-172-1	Sequence 1, Appli
	33	102	2.7	4310	4	US-09-210-361-5	Sequence 5, Appli
	34	102	2.7	4310	4	US-09-740-274-5	Sequence 5, Appli
	35	101.5	2.7	2253	1	US-07-854-596B-39	Sequence 39, Appl
	36	101.5	2.7	3727	1	US-08-249-380-1	Sequence 1, Appli
	37	101	2.7	3668	4	US-09-302-620B-89	Sequence 89, Appl
	38	100.5	2.7	2836	4	US-09-157-257-3	Sequence 3, Appli
	39	100.5	2.7	3278	1	US-08-484-105-13	Sequence 13, Appl
	40	100.5	2.7	3278	1	US-08-484-106-13	Sequence 13, Appl
	41	100.5	2.7	5020	4	US-08-961-527-142	Sequence 142, App
	42	100.5	2.7	8501	4	US-09-298-367B-6	Sequence 6, Appli
	43	100	2.7	4330	3	US-09-310-293-1	Sequence 1, Appli
	44	100	2.7	4330	4	US-09-579-376-1	Sequence 1, Appli
	45	100	2.7	6744	1	US-08-119-125A-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-221-017B-726
; Sequence 726, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER: 32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  726:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1974 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:   1...1974
US-09-221-017B-726

Alignment Scores:
Pred. No.:              1.3e-317      Length:         1974
Score:                  2995.00       Matches:        576
Percent Similarity:     99.65%        Conservative:   0
Best Local Similarity:  99.65%        Mismatches:     0
Query Match:            80.53%        Indels:         2
DB:                     4             Gaps:           0

US-10-008-355-2 (1-712) x US-09-221-017B-726 (1-1974)

Qy    137 AspLysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAla 156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCT 60

Qy    157 GlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIle 176
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATC 120

Qy    177 ValGluProPheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAsp 196
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GTAGAGCCTTTCTATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGAC 180

Qy    197 ValArgMetValPheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsn 216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTTCGTATGGTATTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAAC 240

Qy    217 TrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAsp 236
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGGATGTGGCCGCGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGAC 300

Qy    237 AsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAla 256
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AACCGGCCGGCCGAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCC 360

Qy    257 ValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySer 276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 420

Qy    277 ThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnPro 296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCT 480

Qy    297 ArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGln 316
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CGTATCGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAG 540

Qy    317 AlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSer 336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCTACCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCG 600

Qy    337 IleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGlu 356
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATCGGTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAA 660

Qy    357 ArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeu 376
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AGAGCATTCGCAGACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTG 720

Qy    377 SerSerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeu 396
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TCTTCTCTCGAAAAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTG 780

Qy    397 SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeu 416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTG 840

Qy    417 AlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAsp 436
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GCTACAAATCCTGATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGAC 900

Qy    437 TyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArg 456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TACCTCCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGT 960

Qy    457 IleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGly 476
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 ATCCCTGCCGACAAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC 1020

Qy    477 AspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAspLys 496
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GACACGAAGAAGTATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAG 1080

Qy    497 PheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAsp 516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TTCCATGCCATGCTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGAT 1140

Qy    517 ProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAla 536
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CCGGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCG 1200

Qy    537 MetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGlu 556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ATGGCCAATGCCTATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAG 1260

Qy    557 MetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGly 576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGC 1320

Qy    577 SerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLys 596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TCCATCAAGGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAG 1380

Qy    597 GlyValLeuGluLysGlnAsp-Pro-LysSerAspGluPheAlaValGlnGluAsnIleL 616
          ||||||||||||||||||||| |||  |||||||||||||||||||||||||||||||||
Db   1381 GGCGTATTGGAGAAGCAGGATTCCTTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCC 1440

Qy    616 euAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleA 636
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TCGACCTCTTCCGCACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCG 1500

Qy    636 laPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysA 656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CTTTCCTATCGAACAACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGA 1560

Qy    656 snGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleG 676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ACGGCCGTCTGATCGGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCG 1620

Qy    676 luPheGluProAspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetI 696
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AGTTCGAACCCGATCTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGA 1680
```

```
Qy     696 leAspLysTrpGlyGlnCysProArgLeuIleGlnGluLeuLysLeuIle 712
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 TTGACAAATGGGGTCAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATC 1730
```

RESULT 2
US-09-221-017B-382
; Sequence 382, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 382:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1317 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...1317
US-09-221-017B-382

Alignment Scores:

Pred. No.:	1.07e-71	Length:	1317
Score:	743.50	Matches:	168
Percent Similarity:	55.38%	Conservative:	74
Best Local Similarity:	38.44%	Mismatches:	172
Query Match:	19.99%	Indels:	23
DB:	4	Gaps:	7

US-10-008-355-2 (1-712) x US-09-221-017B-382 (1-1317)

```
Qy      41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60
           :::|||:::|||  |||        :::         ||||||:::   :::
Db      16 CAGATGAAAGAGCGCGGCCTTAAGATGAAGGAATACGACCTTTATAATCCCAACGGCACA 75

Qy      61 SerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80
           |||:::   :::|||||||||:::|||  ||||||||||||||    |||||||||
Db      76 TCGCTGAAAGATGCCGTAGTGCTATTCGACGGAGGATGTACGGGCGAGGTCGTTTCCGAT 135

Qy      81 GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100
           :::||||||:::   ||||||||||||||||||||         |||||||:::   ||||||
Db     136 CGCGGACTGGTACTGACCAATCACCACTGCGGATACGATATGATCCAGGCTCACAGCACG 195

Qy     101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuPro 120
           ::::::|||:::||||||   :::||||||   :::                 :::|||||||||
Db     196 CTCGAGCATAACTATCTCGAAAATGGATTTTGGGCGATGAGAGAAGCGGATGAATTACCG 255

Qy     121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140
              :::||||||   ::::::   ||||||        |||||||||   |||:::
Db     256 AACAAGGATATTTCCGTGGTATTCATCGACAAGATCGAAGATGTCACAGACTACGTCAAG 315

Qy     141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160
                 ||||||   |||   |||                          :::   :::
Db     316 AAAGATCTCAAAGCCATCAAAGATCCCAACAGCATGGACTACCTCTCTCCGAAGTACCTG 375

Qy     161 GlnGluLeuAlaLysLysGlu---------------AsnAlaAspGluAsnGlnLeuCys 175
           |||:::||||||   |||:::                  :::|||               |||
Db     376 CAAAAGTTGGCTGACAAGAAGGCCGGCAAGAACTTTTCTGCCAAGAATCCGGGGCTTTCC 435

Qy     176 Ile---ValGluProPheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPhe 194
           :::   ::::::   ||||||         |||   |||   :::               :::
Db     436 GTAGAGATCAAAGCCTTCTATGGGGGCAATCTCTACCTGATGTTTACCAAAAAGACTTAT 495

Qy     195 LysAspValArgMetValPheAlaProProSerSerValGlyLysPheGlyGlyAspThr 214
              |||||||||:::|||   |||||||||:::|||:::||||||||||||   ||||||
Db     496 ACGGATGTTCGACTGGTGGGAGCACCTCCCACCAGCATTGGCAAATTCGGTGCCGATACG 555

Qy     215 AspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGly 234
           |||||||||:::|||||||||||||||||||||||||||:::||||||:::||||||
Db     556 GACAACTGGATCTGGCCTCGTCATACTGGCGACTTCTCCATCTTCCGTATCTATGCGGAC 615

Qy     235 AlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPhe 254
              :::         ||||||   ||||||:::||||||   |||   ||||||         |||
Db     616 AAGAATGGCAATCCCGCACCATACTCTGAAGATAATGTTCCGCTCAAGCCGAAGCGTTTC 675

Qy     255 AlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPhePro 274
              :::|||:::   |||   :::   :::|||||||||||||   :::|||||||||
Db     676 TTCAATATCTCCCTTGGTGGAGTACAAGAGAACGACTACGCCATGATAATGGGTTTCCCC 735

Qy     275 GlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsn 294
           |||:::|||   ||||||   |||:::         |||::::::            :::|||
Db     736 GGTACTACGCACCGCTATTTCACGGCTTCCGAAGTAGACGAATGGAAAAGCATCGACAAC 795

Qy     295 AsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAla 314
           :::   ||||||   :::|||   |||:::||||||:::               |||   |||
Db     796 GATATTCGCATCCGCATGCGTGATATTCGTCAGGGTGTCATGCTCAGGGAAATGCTGGCC 855

Qy     315 AspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLys 334
           |||         :::|||   |||:::::::||||||||   |||   |||   :::|||
Db     856 GATCCTCAGATCAAAATCATGTATTCAGCTAAATATGCCGCTTCGCAGAATGCTTACAAA 915

Qy     335 AsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAla 354
           :::||||||   |||         :::                :::         |||:::|||
Db     916 CGTGCTATAGGTGCCAACTGGGCGATCAAGACACGCGGCCTGCGTCAAAACAAACAGGCG 975

Qy     355 GluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAsp 374
              :::                   |||   |||   |||   :::         |||   :::
Db     976 ATGCAGGACAGGCTGATCGCATGGGGAGCGAAGCAGGGT---ACTCCTCGATATGAAGAG 1032

Qy     375 ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThr 394
              :::      ::::::               :::   |||         |||
Db    1033 GCCGTACACGAAATCGATGCTACGGTAGCCAAGCGTGCGGATCTGCGCCGTCGTTATTGG 1092
```

```
Qy    395 TyrLeuSerGluThrLeuPheGlyGlyThrGluValValArg-----------PheAla 410
             :::    |||    :::        |||    |||       |||
Db   1093 ATGATAGAAGAAGGCATCATCCGTGGGATCGAGTTTGCCCGTTCTCCTATCCCCACCGAA 1152

Qy    411 GlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeu 430
                   ||||||      |||   ||||||   |||           |||
Db   1153 GATGAGACGAAAGCTCTGCAAGGCAAT---GATGCTTCGGCT------CGCAAGGAGGCG 1203

Qy    431 AspAspLysTyr--------------------------LysAspTyrLeuProSerLeu 441
             ||||||                               |||||||||         :::
Db   1204 ATCGATAAGATTCGTACACGCTACAGCAAATTTGCCAACAAGGACTACAGTGCAGAGGTG 1263

Qy    442 AspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIlePro 458
           |||:::||||||       |||||||||          :::    :::|||
Db   1264 GACAAGAAAGTGGCCGTAGCCATGCTGACGGAATACCTCAAGGAAATGCCC 1314

RESULT 3
US-09-221-017B-1045
; Sequence 1045, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  1045:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2384 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...2384
US-09-221-017B-1045

Alignment Scores:
Pred. No.:               2.82e-38      Length:         2384
Score:                   441.50        Matches:        90
Percent Similarity:      61.84%        Conservative:   38
Best Local Similarity:   43.48%        Mismatches:     74
Query Match:             11.87%        Indels:         5
DB:                      4             Gaps:           2

US-10-008-355-2 (1-712) x US-09-221-017B-1045 (1-2384)

Qy    507 GluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLysSerValIle 526
          ||||||   |||              :::|||||||   |||        :::   ||||||
Db     18 GAAAAGCTTGCC-----------GAAGATCCTATGGTACTCTTCGCCTCTTCGGTTTTC 65

Qy    527 AlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGly 546
                   |||    :::.    :::                 |||    :::
Db     66 GACGAATACCGCAAACTCTACAACGAACTTCGTCCCTATGACGATCCTATTCTAAGGGCA 125

Qy    547 LysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAsp 566
          :::|||    :::    |||||||||   ||||||             :::          |||
Db    126 CAACGCACCTACATTGCAGGTCTCTTGGAAATGGATGGCGATCAGGATCAATTCCCGGAT 185

Qy    567 AlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGly 586
          ||||||   |||:::|||   :::||||||   :::|||||||||   |||:::|||
Db    186 GCTAACCTGACACTTCGTTTCACCTATGGTCAAGTGAAGGGCTATTCACCCCGTGACAAT 245

Qy    587 AlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSer 606
             :::|||    :::    ||||||       ||||||:::||||||:::||||||    :::
Db    246 GTTTACTACGGACATCAAACCACATTGGATGGTGTGATGGAAAAAGAAGATCCCGATAAT 305

Qy    607 AspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArg 626
             ||||||   |||           :::         ::::::        |||::::::||||||
Db    306 TGGGAATTTGTAGTCGATCCCAAGCTGAAAGCCGTATACGAGCGTAAAGACTTCGGGCGT 365

Qy    627 TyrAlaGlu---AsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly 645
          ||||||:::    :::|||::::::    :::||||||    :::             ||||||
Db    366 TATGCCGATCGCAGCGGTCGCATGCCTGTAGCCTTTTGCGCCACCACACATACAACCGGC 425

Qy    646 GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAsp 665
          ||||||||||||||||||||    :::    ||||||   ||||||||||||   ||||||
Db    426 GGCAACTCAGGCAGTCCGGTCATGAATGCCAACGGCGAACTGATCGGTCTCAACTTCGAT 485

Qy    666 GlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 685
             |||||||||    :::   |||||||||::::::          |||    ||||||:::|||
Db    486 CGTAACTGGGAGGGAGTCGGTGGCGACATCCAGTATCTGGCCGACTACCAGCGCAGCATC 545

Qy    686 SerValAspIleArgTyrValLeuPheMetIleAspLysTrpGlyGlnCysProArgLeu 705
             |||||||||||||||||||||    :::||||||||||    |||    |||   ||||||
Db    546 ATTGTGGATATTCGCTACGTCCTGCTTGTGATAGACAAAGTAGGCGGTTGCCAACGCCTG 605

Qy    706 IleGlnGluLeuLysLeuIle 712
          :::    |||:::    ::::::
Db    606 TTGGATGAAATGAATATCGTT 626

RESULT 4
US-09-221-017B-12
; Sequence 12, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:   1...561
US-09-221-017B-12

Alignment Scores:
Pred. No.:                5.28e-20      Length:          561
Score:                    265.50        Matches:         59
Percent Similarity:       96.83%        Conservative:    2
Best Local Similarity:    93.65%        Mismatches:      2
Query Match:              7.14%         Indels:          1
DB:                       4             Gaps:            0

US-10-008-355-2 (1-712) x US-09-221-017B-12 (1-561)

Qy      1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db    366 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTAAGGG 425

Qy     21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 485

Qy     41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60
          ||||||||||||||||||||||||   |||:::|||||||||||||||||||||||||||
Db    486 CGAATGCGTGAGCTCGGCTTTACCT--CCCGTGGATTCGCTCTACAGTTTCGACAAGCCG 543

Qy     61 SerIleAla 63
          |||:::|||
Db    544 TCCATGGCA 552

RESULT 5
US-08-953-492-1
; Sequence 1, Application US/08953492
; Patent No. 5849555
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, James
;    APPLICANT:  Jaworski, Deborah
;    APPLICANT:  Lawlor, Elizabeth
;    APPLICANT:  Wang, Min
;    TITLE OF INVENTION:  NOVEL valS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/953,492
;      FILING DATE:  17-OCT-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/844,064
;      FILING DATE:  18-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9607991.8
;      FILING DATE:  18-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P31458-4/1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2652 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-953-492-1

Alignment Scores:
Pred. No.:                0.000381      Length:          2652
Score:                    130.00        Matches:         106
Percent Similarity:       32.81%        Conservative:    62
Best Local Similarity:    20.70%        Mismatches:      166
Query Match:              3.50%         Indels:          178
DB:                       2             Gaps:            25

US-10-008-355-2 (1-712) x US-08-953-492-1 (1-2652)

Qy    192 AspValPhe------------------------LysAspValArgMetValPheAlaPro 203
          |||||||||                        |||            :::|||   |||
Db     76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 135

Qy    204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
          |||:::      ||||||                                    |||
Db    136 CCAAACGTTACAGGTAAACTT---------------------------------CACCTT 162

Qy    224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
          |||                                                   ::::::
Db    163 GGT---------------------------------------------------CACGCT 171

Qy    244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
             |||            :::   :::                  |||||||||:::
Db    172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAAACGCATGCAAGGTTTTGAT 228
```

```
Qy     264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                                   |||:::      |||||            :::   |||
Db     229 -----------------ACCCTTTGGCTTCCTGGGATGGACCACGCAGGGATTGCCACT 270

Qy     283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                   ||||||:::|||:::          |||         |||    ::::::
Db     271 CAGGCTAAGGTAGAGGAGCGCTTGCGTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT 330

Qy     303 -----------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                                |||        ||||||:::            :::   ||||||
Db     331 GAGTCTTTCTTGACGAAAGTCTGGGAATGGAAAGAC-----------GAATATGCCACT 378

Qy     319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
              ||||||             |||      |||   :::|||   :::
Db     379 ACTATCAAGGAACAATGGGGCAAGATGGGGCTCTCTGTAGACTATTCTCGTGAGCGTTTC 438

Qy     338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
              ::::::   |||||||:::                           :::|||      :::
Db     439 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 474

Qy     358 AlaPheAlaAspTrpIleArgLys-----------AsnGlyLysSerAlaValTyrGly 373
              |||   |||      :::|||                 |||:::         :::
Db     475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 534

Qy     374 Asp---------------ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys 388
           |||                   ||||||   :::|||      :::|||:::
Db     535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTGGAAGGT 594

Qy     389 AlaAsnArgGluMetThrTyrLeu------------------------------------ 396
           |||         |||   |||:::
Db     595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGTCCTTGAAGTTGCTACA 654

Qy     397 ------SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                         |||||:::||||||
Db     655 ACTCGTCCTGAGACCATGTTTGGGGACGTT------------------------------ 684

Qy     415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
           |||:::|||   ||||||:::                            |||   :::|||
Db     685 GCGGTTGCGGTCAACCCAGAA--------------------------GACCCGCGCTAC 717

Qy     435 LysAsp-----------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
           ||||||                      ||||||   :::      |||::::::|||
Db     718 AAGGACTTGATTGGTAAAAATGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy     449 MetLeuAsp---------------------------------------IleValArgArg 455
           :::   |||                                        |||
Db     775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy     456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
                 |||   |||   |||               |||:::
Db     835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGTCATG 894

Qy     476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
              |||               |||      |||:::   |||   :::            |||
Db     895 AACGACGACGGAACCATGAATGACTTGGCCTTTGAATTTTCAGGCATG---------GAC 945

Qy     496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
           :::|||   |||                  |||         ||||||      |||
Db     946 CGTTTTGAAGCTCGT--------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy     516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----------- 531
                 |||::::::   |||   |||   :::      :::
Db     991 GGTGCCCTCGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAGCGTACAGGT 1050

Qy     532 -------------------------------------------------IleGlnAla 534
                                                            :::
Db    1051 GTTGTGGTTGAACCTCGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy     535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
           :::|||:::||||||          :::|||
Db    1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG-------------------------- 1143

Qy     555 ArgGluMetTyrProGlyArgAlaLeuProSer-------------------------- 565
              |||   ||||||   |||              :::
Db    1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203

Qy     566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
              |||         :::   ::::::   :::|||                   |||
Db    1204 GACTGGGTTATCTCTCGTCAGCTCTGGTGGGGTCAC-----------------CAAATC 1245

Qy     586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
              ||||||||||||            |||:::   :::   ::::::   |||:::
Db    1246 CCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy     606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
              |||   :::         |||::::::||||||
Db    1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335
```

RESULT 6
US-08-961-527-27/c
; Sequence 27, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
;   APPLICANT: Charles Kunsch
;   TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
;   NUMBER OF SEQUENCES: 391
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: Maryland
;   COUNTRY: USA
;   ZIP: 20850
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2
;   SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/961,527
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Brookes, A. Anders
;   REGISTRATION NUMBER: 36,373
;   REFERENCE/DOCKET NUMBER: PB340P1
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 309-8504
;   TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 27:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 6004 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
US-08-961-527-27

```
Alignment Scores:
Pred. No.:                0.00173      Length:          6004
Score:                    129.00       Matches:         116
Percent Similarity:       33.62%       Conservative:    77
Best Local Similarity:    20.21%       Mismatches:      194
Query Match:              3.47%        Indels:          187
DB:                       4            Gaps:            26

US-10-008-355-2 (1-712) x US-08-961-527-27 (1-6004)

Qy     131 LysIleValLysValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMet 150
           ||||||                              |||         |||   |||   ::::::::
Db    4315 AAAATAAATAATACGAATTTGTCTGCAGAAGAAGCTGCTAAGCAAATTCAAAATAAAATA 4256

Qy     151 GluArgLeu-ArgLysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAs 170
              :::::: ||||||          :::   :::||||||:::   |||   |||
```

```
Db   4255 AACAAAATAGAGAAAGGACACACACATGTCTAAAGAACTTTCACCTAAATACAATCCAGC 4196

Qy    170 pGluAsnGlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyrPheLeuIleVa 190
           |||  :::                :::  :::                        :::
Db   4195 CGAGGTTGAGGCTGGTCGTTACCAAAAATGG-----------------------CTTGA 4160

Qy    190 lTyrAspValPhe-----------------------LysAspValArgMetValPheAl 202
              |||||||||                          |||         :::|||
Db   4159 TGCGGATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTGATTCC 4100

Qy    202 aProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHi 222
           ||||||:::        ||||||                                   ||
Db   4099 ACCACCAAACGTAACTGGGAAACTT--------------------------------CA 4073

Qy    222 sThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTy 242
           |   |||                                                   ::
Db   4072 CCTTGGT-----------------------------------------------CA 4064

Qy    242 rSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTy 262
           ::::   |||              :::   :::                 |||||||||::
Db   4063 CGCTTGGGATACG---ACTTTGCAGGATATCATCATCCGTCAAAAACGCATGCAAGGCTT 4007

Qy    262 rLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu-- 281
           :                       |||:::   ||||||                :::
Db   4006 TGAT------------------ACCCTTTGGCTTCCAGGTATGGACCACGCTGGTATTGC 3965

Qy    282 -ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValAr 301
           |||        ||||||:::|||:::          |||          |||   ::::::
Db   3964 GACTCAGGCTAAGGTTGAGGAGCGCTTACGTGGTGAGGGTATTACGCGTTATGACCTAGG 3905

Qy    301 gGly-----------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAl 317
                           |||       ||||||:::               :::   ||
Db   3904 TCGTGAGTCTTTCTTGACGAAGGTCTGGGAATGGAAAGAC-----------GAATATGC 3857

Qy    317 aThrArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSe 336
           ||||  ||||||         |||         |||   :::|||   :::
Db   3856 CACTACCATCAAGGAACAATGGGGCAAGATGGGGCTTTCTGTAGACTATTCTCGTGAGCG 3797

Qy    336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGl 356
            . ::::::   |||||||:::                          :::|||
Db   3796 TTTCACTCTTGATGAAGGTTTGTCA-----------------------AAAGCTGTTCG 3761

Qy    356 uArgAlaPheAlaAspTrpIleArgLys-----------AsnGlyLysSerAlaValTy 372
           :::   |||   :::        :::|||              |||:::       :::
Db   3760 TAAGGTCTTTGTGAACCTTTACAAGAAAGGCTGGATCTATCGTGGTGAGTTTATCATCAA 3701

Qy    372 rGlyAsp--------------ValLeuSerSerLeuGluLysAlaTyrLys-------- 384
             |||                 ||||||   :::|||      :::|||
Db   3700 CTGGGACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTAGA 3641

Qy    385 ---------------------------------GluGlyAlaLysAlaAsnArgGluMe 393
                                            :::|||:::::|||
Db   3640 AGGTGCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGCCCTTGAAGTTGC 3581

Qy    393 tThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAl 413
           |||         ||||||:::||||||
Db   3580 TACAACTCGTCCTGAGACTATGTTTGGGGACGTT------------------------- 3547

Qy    413 aAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLy 433
             |||:::|||   ||||||:::                        |||    ::
Db   3546 ----GCGGTTGCGGTTAATCCAGAA-------------------------GACCCGCG 3518

Qy    433 sTyrLysAsp-----------------TyrLeuProSerLeuAspArgLysValLeuPr 447
           :|||||||||                  ||||||   :::       |||::::::|||
Db   3517 CTACAAGGACTTGATTGGTAAAAACGTCATCCTTCCA---ATCGCTAATAAACTCATCCC 3461

Qy    447 oAlaMetLeuAsp--------------------------------------IleValAr 454
           |    :::  |||                                       |||
Db   3460 AATCGTTGGAGATGAGCACGCAGATCCTGAGCTTGGTACTGGTGTCGTGAAAATCACACC 3401

Qy    454 gArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPh 474
                 |||   |||   |||               |||:::
Db   3400 TGCCCACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGT 3341

Qy    474 eLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSe 494
              |||            :::      |||:::   |||    :::
Db   3340 CATGAACGACGACGGAACTATGAATGAGCTTGCCTTTGAATTTTCAGGCATG-------- 3289

Qy    494 rAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGl 514
             |||:::|||   |||                 |||         ||||||       ||
Db   3288 -GATCGTTTTGAAGCTCGT---------------AAGGCAGTCGTTGCTAAGTTGGAAGA 3245

Qy    514 uLysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla-------- 531
           |               |||:::::   |||   |||    :::           :::
Db   3244 AATCGGTGCCCTTGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAACGTAC 3185

Qy    532 -----------------------------------------------------IleGl 533
                                                                  :::
Db   3184 AGGTGTCGTAGTTGAGCCACGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGC 3125

Qy    533 nAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGl 553
             :::|||:::||||||           :::|||
Db   3124 TAAGAACGCCATTGCTAACCAAGACACAGAGGACAAG--------------------- 3088

Qy    553 yLeuArgGluMetTyrProGlyArgAlaLeuProSer---------------------- 565
                |||   ||||||   |||         :::
Db   3087 ----GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGT 3032

Qy    566 ----AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGl 584
                |||         :::   ::::::   :::|||                      ||
Db   3031 CCACGACTGGGTTATCTCTCGTCAGCTTTGGTGGGGTCAC-----------------CA 2990

Qy    584 nAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspPr 604
           |      ||||||||||||           |||:::        :::   ::::::   ||
Db   2989 AATCCCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCC 2936

Qy    604 oLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
           |:::   |||   :::           |||::::::||||||
Db   2935 AGAAGGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 2896
```

RESULT 7

US-09-059-584-52

```
; Sequence 52, Application US/09059584
; Patent No. 6440701
;  GENERAL INFORMATION:
;    APPLICANT:  Myers, Lisa E
;    APPLICANT:  Schryvers, Anthony B
;    APPLICANT:  Harkness, Robin E
;    APPLICANT:  Loosmore, Sheena M.
;    APPLICANT:  Du, Run-Pan
;    APPLICANT:  Yang, Yan-Ping
;    APPLICANT:  Klein, Michel H
;    TITLE OF INVENTION:  Transferrin Receptor Genes of Moraxella
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6th Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/059,584
;      FILING DATE:  14-APR-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/778,570
;      FILING DATE:  03-JAN-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I
```

```
;      REGISTRATION NUMBER:  24973
;      REFERENCE/DOCKET NUMBER:  1038-794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2142 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-059-584-52

Alignment Scores:
Pred. No.:                 0.00141      Length:          2142
Score:                     123.50       Matches:         125
Percent Similarity:        32.17%       Conservative:    77
Best Local Similarity:     19.90%       Mismatches:      273
Query Match:               3.32%        Indels:          153
DB:                        4            Gaps:            25

US-10-008-355-2 (1-712) x US-09-059-584-52 (1-2142)

Qy     119 LeuProIleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLys 138
            :::|||              ::::::|||||||:::|||          :::|||     |||
Db     466 ATTCCTTTTGATAAAAATCTTATTGAATACCTTAAAAAATCATCCGAGGTTGTAAGTAAA 525

Qy     139 ValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGlu 158
              |||   |||      ||||||    :::                    |||
Db     526 TTTGAAGCACAAAAAGGCGGTATTGAAAATAACACAAGACTGACACACAAAGATTTATCA 585

Qy     159 ValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGlu 178
                    |||:::   ||||||   :::   |||   :::|||   |||
Db     586 TCAGAGCAAAAAGAAGCAAAAGTCAAAGAAGCGTTGGACAATGCTTTA------------ 633

Qy     179 ProPheTyrSerAsnAsnGluTyrPheLeuIleVal---------TyrAspValPheLys 195
                    ::::::      :::|||          ::::::          :::|||
Db     634 ACTCAATTTGCCCAAGAAAAATACAAGGAGCTAATTGAGAACGCCCATGATAAAAAATCT 693

Qy     196 AspValArgMetValPheAlaProProSerSerValGlyLysPheGlyGlyAspThrAsp 215
            |||   |||
Db     694 GACGCACGC--------------------------------------------------- 702

Qy     216 AsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAla 235

Db     702 ------------------------------------------------------------ 702

Qy     236 AspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAla 255
              ||||||      ||||||   |||            |||       |||   |||
Db     703 ---AACCGTGATCTAGAATATGTCAAGTCTGGTTTTAACTATCTTTCTGGATATACCGCC 759

Qy     256 Ala---------------ValSerMetGlnGly------------------TyrLysAla 264
                             :::   :::|||                    ||||||
Db     760 ACCGACCACGACAAAAAAACCAATTATCGTGGCTATTATGGTGCGTTGTATTATAAAGGC 819

Qy     265 AspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrp 284
             :::   |||      :::         |||   ::::::   :::|||         |||
Db     820 AGCGAAACCGCCAAAGAGCTA------CCACAAACAAGTGCAAAATATAAAGGTTATTGG 873

Qy     285 Gly------ValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                    :::      ::::::|||:::                ::::::   |||
Db     874 GACTTTATGACAGATGCCACACTTGATAACAAATACACG---------GATTTGCCAGGT 924

Qy     303 Ile---LysGlnGlyIleTrpLysGluAlaMetSerAlaAspGln---AlaThrArgIle 320
            |||   :::|||      |||:::      :::||| |||:::   ||||||   :::
Db     925 ATCGCCAGACAAACCCAGTGGCGTAGTCTTGTTTCTACTGATGAGTATGCAACGCTCTTG 984

Qy     321 -----------------------LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyr 332
                                  |||         |||      |||:::   :::
Db     985 ACAGACAAAAATAACAAGCCCAGTGATTACAATGGTGCATATGGTCATAGCAGTGAATTT 1044

Qy     333 TrpLysAsnSerIleGly--------------------MetAsnArgGlyLeuAlaArg 345
               |||                                ::::::      |||:::
Db    1045 GATGTTAATTTTGCTGATAAAAAAATTAAAGGCAAACTTATCAGTAATCAGTTATCAGGC 1104

Qy     346 LeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLys 365
               |||      :::::::|||   :::         |||               |||
Db    1105 ACAGCTGTAACCGCCAAAGAGCGTTATAAAATAGAAGCTGATATCCACGGCAACCGCTTC 1164

Qy     366 AsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGlu 385
              |||   :::                     ::::::   :::||||||         |||
Db    1165 CGTGGCAGTGCC--------------------ACCGCAAGCGATAAAGCA------GAA 1197

Qy     386 GlyAlaLysAlaAsnArgGluMetThrTyr---LeuSerGluThrLeuPheGlyGlyThr 404
             :::|||              |||         :::       |||   ||||||
Db    1198 GACAGCAAAACCCAACACCCCTTTACCAGCGATGCTACAAACAAGCTAGAAGGTGGTTTT 1257

Qy     405 GluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla 424
                  :::         :::   |||               |||:::   :::
Db    1258 TATGGACCAAAAGGCGAGGAGCTGGCAGGTAAATTCTTAACCGATGACAACAAACTCTTT 1317

Qy     425 GlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLys 444
            |||:::      :::      ||||||   :::               ||||||
Db    1318 GGGGTCTTTGGTGCTAAACGAGATAAAGTAGAAAAAACCGAAGCCATCTTAGATGCCTAT 1377

Qy     445 ValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeuProAsp 464
              |||
Db    1378 GCACTT------------------------------------------------GGG 1386

Qy     465 IlePheLysAsnValIleAsp---LysLysPheLysGlyAspThrLysLysTyrAlaAsp 483
              |||   |||                  |||         |||||||||      |||
Db    1387 ACATTTAATAATACAAATAAAGCAACCACATTCACCCCATTTACCAAAAAACAACTGGAT 1446

Qy     484 ---PheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLys 502
              |||      |||   :::|||      |||         :::   :::
Db    1447 AACTTTGGCAATGCCAAAAAGTTGGTCTTGGGTTCTACCGTCATTAATTTGGTGTCTACC 1506

Qy     503 SerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer 522
                 |||   :::|||   |||         ||||||
Db    1507 GATGCCACCAAAAAATGAATTCACCAAAAAAATTCACCAAAGAC----------------- 1548

Qy     523 LysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAla 542
            |||         |||      :::|||   :::            |||   |||
Db    1549 AAGCCAACTTCTGCCACAAACAAAGCGGGCGAGACT---TTGATGGTGAATGATGAAGTT 1605

Qy     543 IleGluLys------GlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560
            |||   |||         ||||||            :::         |||:::   |||
Db    1606 ATCGTCAAAACCTATGGCAAAAACTTTGAATACCTAAAATTTGGTGAGCTTAGTGTCGGT 1665

Qy     561 ArgAla------------LeuProSerAspAlaAsnPheThrMetArgMetSerTyrGly 576
             :::                  |||      :::      |||
Db    1666 GATAGCCATAGCGTCTTTTTACAAGGCGAACGCACCGCTACCACAGGCGAGAAAGCCGTA 1725

Qy     577 SerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThr--------- 593
                |||                  |||   |||      |||   |||
Db    1726 CCAACCACAGGCAAAGCCAAATATCTGGGGAACTGGGTAGGATACATCACAGGAGCGGGC 1785

Qy     594 ThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGlu 613
            ||||||||            |||   ||||||                    :::   :::
Db    1786 ACAGGAAAAAGCTTTAATGAGGCCCAAGAT--------------------ATTGCTGAT 1824

Qy     614 AsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeu 633
               :::|||      |||               |||:::               |||:::
Db    1825 TTTGACATTGACTTTGAGAGAAAATCAGTTAAAGGCAAACTGACCACCCAAGGCCGCACA 1884

Qy     634 HisIleAlaPheLeuSerAsnAsnAspIleThrGly------------------------ 645
               |||         :::|||   |||
Db    1885 GATCCTGTCTTTAACATCAAAGGTGAAATTGCAGGCAATGGCTGGACAGGCAAAGCCAGC 1944

Qy     646 ----------------------------------------GlyAsnSerGlySerProVal 652
                                                   ::::::|||         :::
Db    1945 ACCACCAAAGCGGACGCAGGAGGCTACAAGATAGATTCTAGCAGTACAGGCAAATCCATC 2004

Qy     653 PheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSer 672
               :::|||      :::   |||         :::      |||            |||
Db    2005 GTCATCGAAAATGCCGAAGTTACTGGGGGCTTTTATGGTCCAAATGCAAACGAGATGGGC 2064
```

```
Qy     673 GlyAspIleGluPheGluProAsp 680
           |||          :::   |||
Db    2065 GGGTCATTTACACACGATACCGAT 2088

RESULT 8
US-08-793-824-1
; Sequence 1, Application US/08793824
; Patent No. 5981838
;  GENERAL INFORMATION:
;    APPLICANT:  Simpson, Christine Lynn
;    APPLICANT:  Giffard, Philip Morrison
;    APPLICANT:  Jacques, Nicholas Anthony
;    TITLE OF INVENTION:  Genetic Manipulation of Plants to
;    TITLE OF INVENTION:  Increase Stored Carbohydrates
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Griffith Hack & Co
;      STREET:  Level 8, 168 Walker Street
;      CITY:  No. 5981838th Sydney
;      STATE:  New South Wales
;      COUNTRY:  Australia
;      ZIP:  2060
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/793,824
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PM7643
;      FILING DATE:  24-AUG-1994
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  61 2 9957 5944
;      TELEFAX:  61 2 957 6288
;      TELEX:  26547
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4853 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptococcus salivarius
US-08-793-824-1

Alignment Scores:
Pred. No.:               0.0423      Length:         4853
Score:                   115.00      Matches:        161
Percent Similarity:      34.10%      Conservative:   106
Best Local Similarity:   20.56%      Mismatches:     262
Query Match:             3.09%       Indels:         257
DB:                      2           Gaps:           40

US-10-008-355-2 (1-712) x US-08-793-824-1 (1-4853)

Qy      70 GlyGlyGlyCysThrGlyIleThrValSerAsp---------------GlnGlyLeuIle 84
            ||||||   |||            :::::::::|||          ||||||   |||
Db    1776 GGTGGACAATGTGGATGCGGATCTCTTGCAGATTGCTTCTGATTACTTCAAGGAAAAATA 1835

Qy      85 PheThrAsnHis---------------HisCysGlyTyr-GlyAlaIleGlnSerGlnSe 99
                                      ||||||   :::       :::::::::    ||
Db    1836 CCGTGTCGCAGATAATGAAGCAAATGCCATTGCTCATTTGTCAATCCTTGAAGCTTGGTC 1895

Qy      99 rThrValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLe 119
            |    ||||||   |||   :::|||                |||    |||    :::||
Db    1896 ATACAACGACCACCAATACAACAAGGAT---------------ACCAAGGGTGCTCAGTT 1940

Qy     119 uProIleProGly---------------LeuSerValLysTyrLeuArgLys-------- 131
            |   |||                    |||       :::|||||||||
Db    1941 GTCTATCGACAATCCACTACGTGAAACGCTTTTGACGACTTTCTTGCGTAAGAGCAATTA 2000

Qy     131 ------------------------------------------------------------ 131

Db    2001 CCGTGGCAGCTTGGAGCGCGTGATTACTAACTCTCTTAACAATCGTTCAAGTGAGCAGAA 2060

Qy     132 ---------------------------IleValLysValThrAsp--------------- 137
                                      |||:::      |||
Db    2061 ACACACGCCACGTGATGCCAATTATATTTCGTGCGAGCCCATGATAGTGAAGTGCAAGC 2120

Qy     138 ------------------------LysValGluGlyGlnLeuLysGlyIleThrAspGl 149
                                   ::::::      :::      |||   |||
Db    2121 TGTTTTGGCTAATATCATCAGCAAGCAGATTAATCCAAAAACAGATGGTTTCACCTTCAC 2180

Qy     149 uMetGluArgLeuArgLysAlaGlnGluValCysGln---GluLeuAlaLys------Ly 166
             |||:::   |||:::::::|||   |||:::      ::::::||||||       ||
Db    2181 TATGGATGAGCTCAAACAGGCCTTCGAAATCTACAATGCGGACATCGCGAAGGCTGATAA 2240

Qy     166 sGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyrSerAsnAsnGluTy 186
            |:::               |||                     :::|||   :::
Db    2241 AAAGTACACCCAGTACAATATCCCAGCTGCCTATGCCACAATGTTGACCAACAAGGATAG 2300

Qy     186 rPheLeuIleValTyr------AspValPheLysAspValArgMetValPheAlaProPr 204
                      ||||||      |||:::|||   |||              |||
Db    2301 TATCACTCGTGTTTACTACGGGGACCTCTTTACCGACGATGGCCAATACATGGCTGAAAA 2360

Qy     204 oSer-------------------------------------------------------- 205
             |||
Db    2361 ATCACCGTACTATAATGCCATCGATGCCCTGCTCCGTGCTCGCATCAAATATGTAGCAGG 2420

Qy     206 ---------------------------------------------------------SerVa 207

Db    2421 TGGTCAAGACATGAAGGTTACTAAGCTTAATGGCTATGAAATCATGTCATCTGTGCGTTA 2480

Qy     207 lGlyLysPheGlyGlyAspThrAspAsnTrpMetTrp------------------ProAr 221
             ||||||   |||   :::      |||      |||                  |||||
Db    2481 TGGTAAG---GGGGCAGAAGAAGCCAACCA-GCTTGGTACTGCTGAAACACGCAACCAAG 2536

Qy     221 gHisThrGlyAspPheSer---ValPheArg-ValTyrAlaGlyAlaAspAsnArg---- 238
            |:::   |||   :::|||          ||| :::      ||||||:::::::|||
Db    2537 GAATGCTGGTCCTTACAGCCAACCGTCCAGATATGAAGTTGGGAGCTAATGATCGTCTGG 2596

Qy     239 -----ProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAla- 256
                           :::         |||:::   |||:::|||:::      :::
Db    2597 TAGTCAATATGGGAGCTGCCCACAAAAATCAGGCTTACCGTCCATTGCTTCTCAGCAAAT 2656

Qy     257 -----------------------------------------------------------V 257

Db    2657 CGACAGGTCTTGCGACCTACCTCAAGGATTCTGATGTACCAGCTGGATTGGTTCGCTATA 2716

Qy     257 alSerMetGlnGly---------TyrLysAlaAspAspTyrAla------MetThrIleG 272
                  ||||||          :::   |||||||||   |||        |||:::
Db    2717 CGGACAATCAAGGGAACTTGACCTTCACGGCAGATGATATTGCTGGTCATTCAACCGTTG 2776

Qy     272 lyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluA 292
                  |||            ||||||      |||                   ||||
Db    2777 AAGTTTCAGGT-----------TATTTGGCAGTTTGGGTGCCAGTGGGTGCCTCAGAAA 2824

Qy     292 snGluAsnAsnProArgIleGluValArgGlyIleLysGlnGly------IleTrpLysG 310
            ||:::   :::   |||   :::                |||:::|||      :::::::::
Db    2825 ACCAA---GATGCCCGAACCAAGGCTTCGAGCACCAAGAAGGGTGAGCAGGTCTTTGAAT 2881

Qy     310 luAlaMetSerAlaAspGlnAlaThr-----------------------------ArgI 320
             :::   :::   |||                                        :
Db    2882 CATCAGCAGCTCTTGATTCACAAGTCATTTACGAAGGCTTCTCAAACTTCCAAGATTTCG 2941

Qy     320 leLysTyrAlaSerLysTyr---------------AlaGlnSerAlaAsnTyrTrpLysA 335
            ::|||      |||:::|||               ||||||:::|||         :::|||
Db    2942 TTAAGACACCAAGTCAGTACACCAACCGTGTTATTGCTCAAAATGCCAAACTCTTCAAA- 3000
```

```
Qy     335 snSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArg---LysArgA 354
                                   |||:::            :::          :::     :
Db    3001 --------------GAGTGGGGAATCACTTCCTTTGAATTTGCGCCACAGTATGTGTCTA 3046

Qy     354 laGluGluArgAlaPheAlaAspTrpIleArgLysAsnGly------------------ 367
           :::::::        |||   |||    |||   :::||||||
Db    3047 GCCAAGACGGCACTTTCTTGGATTCTATCATTGAAAATGGCTACGCCTTCGAGGACCGCT 3106

Qy     368 -----------------LysSerAlaValTyrGly---------AspValLeuSerSerL 379
                            |||:::     ||||||        |||::::::    :::|
Db    3107 ACGATATTGCCATGAGCAAGAACAACAAATACGGTTCACTCAAAGACCTCATGGATGCCC 3166

Qy     379 euGluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluT 399
            ||        :::   ||||||     |||      :::     ::::::   :::
Db    3167 TTCGTGCCCTTCACGCAGAAGGTATCTCAGCCATTGCTGAC------TGGGTTCCAGATC 3220

Qy     399 hrLeuPheGly------GlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuA 417
            ::::::        |||  |||||||||      |||        |||:::
Db    3221 AAATTTACAATCTCCCTGGAAAAGAAGTGGTAACA---GCTTCTCGTACCAATAGCTACG 3277

Qy     417 laThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeu---AspAspLysTyrLysA 436
             |||  |||    :::|||   |||      ||||||          |||   :::
Db    3278 GTACA---CCACGTCCAAATGCTGAAATCTACAATAGCCTCTACGCAGCTAAAACACGTA 3334

Qy     436 spTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgA 456
             :::                  |||          |||   ||||||  ::::::   :
Db    3335 CCTTTGGAAATGACTTCCAAGGCAAGTACGGTGGTGCCTTCCTTGATGAATTGAAGGCAA 3394

Qy     456 rgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysG 476
           ::  ||||||                 |||||||:::   |||   |||
Db    3395 AATACCCAGCA---------------ATCTTTGAGCGCGTGCA-GAT------------- 3425

Qy     476 lyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAspL 496
                                        |||   :::      :::   |||        :
Db    3426 ---------------------------TTCAAACGGCCGTAAATTGACTACCAATGAG 3456

Qy     496 ysPheHisAlaMetLeuLys--SerMetAspLysGluLysPheAlaLysAlaIleGluLy 515
           ::    |||||||||:::      |||:::          |||      :::|||    :::|||
Db    3457 AAAATCACGCAATGGTCAGCCAAGTATTTCAATGGAAGCAATATCCAAGGTACTGGAG-- 3514

Qy     515 sAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAs 535
                                                     |||      :::|||    ::
Db    3515 -------------------------------------CTCGCTAT-GTCCTACAAGATAA 3536

Qy     535 pAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuAr 555
            :|||            ::::::::::::   |||:::                    :::
Db    3537 CGCTACCAACCAATACTTCAGCGTCAAAGCAGGTCAAACCTTCCTTCCTAAACAAATGAC 3596

Qy     555 gGluMetTyrProGly-----ArgAlaLeuProSerAspAlaAsnPheThrMetArgMe 573
            |||:::                |||   :::      |||                    ::
Db    3597 TGAAATTACTGGAAGTGGTTTCCGTAGGGTTGGAGATGAT-----------------GT 3638

Qy     573 tSerTyrGlySerIleLysGlyTyrGluPro---------------GlnAspGlyAla-- 587
           :   |||   ||||||   ||||||                       |||   ||||||
Db    3639 CCAATACCTCTCAATTGGTGGCTACCTTGCTAAGAATACCTTTATTCAAGTCGGTGCCAA 3698

Qy     588 ----TrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSe 606
                ||||||   :::            |||      |||      :::|||
Db    3699 CCAGTGGTATTACTTTGATAAGAATGGCAACATGGTCACAGGTGAGCAGGTCATTGATGG 3758

Qy     606 rAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyAr 626
               ::::::          :::|||   |||   |||            ::::::      ||
Db    3759 CAAGAAATACTTCTTCCTAGACAATGGTCTCCAGCTA---------CGTCATGTCCTTCG 3809

Qy     626 gTyrAlaGluAsnGlyGlnLeuHisIle--------------------AlaPheLeuSe 639
            |        :::|||   ::::::                          ||||||
Db    3810 CCAAGGTAGTGATGGTCATGTGTATTATTACGATCCTAAAGGGGTTCAGGCCTTTAACGG 3869

Qy     639 rAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLe 659
                   |||         |||                    ||||||   ||||||::::::
Db    3870 ATTTTATGATTTTGCGGGTCCTCGCCAAGACGTTCGCTACTTTGATGGCAACGGTCAAAT 3929

Qy     659 u 659
                  :
Db    3930 G 3930
```

RESULT 9
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```
Alignment Scores:
Pred. No.:               4.32e+03      Length:        4411529
Score:                   111.00        Matches:       137
Percent Similarity:      29.45%        Conservative:  60
Best Local Similarity:   20.48%        Mismatches:    187
Query Match:             2.98%         Indels:        286
DB:                      4             Gaps:          35

US-10-008-355-2 (1-712) x US-09-103-840A-1 (1-4411529)

Qy        75 GlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAla 94
               |||   ||||||      |||||||||                 |||   |||:::
Db   4023701 GGTGCCACGGTG------CAGGGCCTA-----------CCGCATATCGGGCAT------ 4023666

Qy        95 IleGlnSerGlnSerThrValAspHisAspTyrLeuArgAspGlyPheValSerArgThr 114
                      :::|||   |||         |||   ||||||            :::::|||
Db   4023665 ------GTCCGCAGCGGGGTCGCCTTCGACATCCTGCGCCGATGGTTGCTCGCGCGT--- 4023615

Qy       115 MetGlyGluGluLeuProIleProGlyLeuSerValLysTyrLeuArgLysIleValLys 134
                                      |||      |||   ::::::|||      :::
Db   4023614 -----------------------GGCTACGACGTCGCGTTTATCCGCAACGTGACCGAC 4023579

Qy       135 ValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArg 154
                :::   ||||||:::
Db   4023578 ATCGAAGACAAGATC--------------------------------------------- 4023564

Qy       155 LysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeu 174
                                      |||||||||         |||
Db   4023563 -----------------------CTGGCCAAGGCCGCCGCGGCGGGC------------ 4023540

Qy       175 CysIleValGluProPheTyr-----------SerAsnAsnGluTyrPheLeuIleVal 190
                       |||::::::               ::::::         |||
Db   4023539 ---------CGGCCGTGGTGGGAGTGGGCGGCTACCCACGAGCGTGCCTTCACCGCGGCC 4023489

Qy       191 TyrAspValPheLysAspValArgMetValPheAlaProProSerSerValGlyLysPhe 210
             ||||||         ||||||               |||||||||:::
Db   4023488 TACGACGCTCTG---GACGTC------------TTGCCGCCGTCCGCGGAG--------- 4023453

Qy       211 GlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe--- 229
                                      ||||||  ||||||          :::
Db   4023452 ---------------------------CCGCGCGCCACCGGGCATATCACCCAGATGATC 4023420

Qy       230 ----------------------------ArgValTyrAlaGlyAlaAspAsnArgPro 239
                                             |||   |||         :::
Db   4023419 GAGATGATCGAGCGCCTGATCCAGGCCGGCCACGCGTATACCGGTGGCGGTGAC------ 4023366

Qy       240 AlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSer--- 258
```

```
                                               |||||||||      :::|||
Db 4023365 -----------------------------------GTTTACTTCGACGTTCTGAGCTAC 4023342

Qy     259 --------------MetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPhe 273
                          :::   |||:::|||  ||||||            |||
Db 4023341 CCGGAGTACGGCCAGTTATCCGGTCACAAGATCGATGACGTCCATCAGGGCGAAGGTGTG 4023282

Qy     274 ---ProGlySerThrAsp---ArgTyrLeuThrSerTrpGlyValGlu-AspArgIleGl 291
                |||      |||  |||      |||  |||      |||   |||:::
Db 4023281 GCTGCCGGCAAGCGGGACCAGCGCGACTTCACTTTGTGGAAGGGCGAAAAGCCGGGTG-- 4023224

Qy     291 uAsnGluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAl 311
            |||         |||        |||::::::  |||||||||
Db 4023223 -AACCGTCGTGGCCGACGCCGTGGGGCCGCGGGCGTCCGGGCTGGCATCTGG-------- 4023173

Qy     311 aMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAs 331
                                                                :::|||
Db 4023172 ---------------------------------------------------AATGCTCG 4023165

Qy     331 nTyrTrp-----------------LysAsnSerIleGlyMetAsnArgGlyLeuAlaAr 345
              |||                  :::|||||||||   :::        |||:::
Db 4023164 GCAATGGCTCGCAGCTATCTCGGGCCGGAATTCGATATCCATTGC-GGTGGAATGGATTT 4023106

Qy     345 gLeu---------------AspValIleGlyArgLysArgAlaGluGluArgAlaPheAl 360
            :::                :::   |||  :::   ||||||      |||||
Db 4023105 AGTCTTCCCGCATCATGAGAACGAAATCGCCCAGAGCCGCGCCGCCGGTGATGGGTTCGC 4023046

Qy     360 a-----------------AspTrpIleArgLysAsnGlyLys------SerAlaValTy 372
            |                 |||:::         |||:::        |||
Db 4023045 CCGCTACTGGCTGCACAATGGCTGGGTGACGATGGGCGGGGAGAAGATGAGCAAGTCGCT 4022986

Qy     372 rGlyAspValLeuSer--------------------------SerLeuGluLysAlaTy 383
            |||:::|||||||||                              |||       ||
Db 4022985 GGGCAACGTGTTGTCCATGCCGGCGATGTTGCAGCGGGTGCGGCCGGCTGAGCTGCGTTA 4022926

Qy     383 rLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGl 403
            |    |||:::         |||   |||          |||||||||
Db 4022925 TTACCTGGGCAGCGCGCACTACCGGTCGATGCTCGAGTTCTCCGAGACC----------- 4022877

Qy     403 yThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHi 423

Db 4022877 ---------------------------------------------------------- 4022877

Qy     423 sAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspAr 443
                            ::::::   |||        |||   |||:::
Db 4022876 ---------------GCTATGCAGGATGCGGTGAAAGCCTATGTCGGG----------- 4022844

Qy     443 gLysValLeuProAlaMetLeuAspIleValArgArgArgIle-----------ProAl 459
               |||        |||     ||||||  |||:::                  |||
Db 4022843 -------CTAGAGGACTTCCTGCACCGCGTTCGCACCCGGGTGGGCGCCGTCTGCCCCGG 4022791

Qy     459 aAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLy 479
            |||    |||
Db 4022790 CGATCCGACCCCA--------------------------------------------- 4022778

Qy     479 sLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAl 499

Db 4022778 ---------------------------------------------------------- 4022778

Qy     499 aMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaVa 519
                                 :::||||||:::|||::::::   |||     :::||
Db 4022777 --------------------------CGGTTCGCCGAAGCGCTCGACGACGACCTGTCCGT 4022743

Qy     519 lGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAs 539
            |   ::::::   :::   |||          ||||||
Db 4022742 TCCGATCGCGCTCGCCGAGATTCACCACGTGCGGGCG----------------------- 4022706

Qy     539 nAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrPr 559
                           :::|||   |||          ||||||      :::
Db 4022705 ---------------GAAGGCAACCGGGCACTCGATGCCGGCGACCACGAC-------- 4022670

Qy     559 oGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLy 579
                  ||||||   |||   :::                |||            |||
Db 4022669 ----GGGGCCTTGCGAAGCGCTAGCGCGATTCGGGCGATGATG---------GGCATCCT 4022623

Qy     579 sGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLe 599
              |||    :::|||   |||        |||
Db 4022622 GGGCTGTGACCCGCTCGACCAGCGCTGG--------------------------------- 4022595

Qy     599 uGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu-- 618
              |||    :::|||       |||        |||        :::   ::::::
Db 4022594 -GAATCCCGAGACGAAACCTCGGCAGCGCTGGCCGCCGTCGATGTGCTGGTCCAGGCTGA 4022536

Qy     619 -PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLe 638
              :::     :::          |||         |||        |||
Db 4022535 ACTACAGAATCGGGAAAAGGCCCGCGAGCAGCGCAACTGGGCGCTC-------------- 4022490

Qy     638 uSerAsnAsnAspIleThrGly------GlyAsnSerGlySerProValPheAspLysAs 656
                  :::::::|||   |||            :::|||        |||     |||
Db 4022489 ----GCCGACGAGATCCGGGGTCGGCTGAAGAGAGCCGGCATCGAGGTCACCGAC----- 4022439

Qy     656 nGlyArgLeuIleGlyLeuAlaPheAspGly---AsnTrpGluAlaMetSerGlyAspIl 675
                                      ||||||      |||        :::    ||||||
Db 4022438 -------------------ACCGCCGACGGCCACAGTGGTCGCTGCTGGGTGGTGACAC 4022398

Qy     675 eGluPheGluProAspLeuGlnArg 683
              :::      |||         |||
Db 4022397 CAAGTAGATGCCCGGTAACTCTCGG 4022373
```

RESULT 10
US-08-961-527-134
; Sequence 134, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-134

Alignment Scores:

Pred. No.:	0.578	Length:	12665
Score:	110.50	Matches:	132
Percent Similarity:	35.47%	Conservative:	100
Best Local Similarity:	20.18%	Mismatches:	246

```
Query Match:            2.97%        Indels:       176
DB:                     4            Gaps:         30

US-10-008-355-2 (1-712) x US-08-961-527-134 (1-12665)

Qy    127 LysTyrLeuArgLysIleVal---------------LysValThrAspLysValGluGly 141
          :::|||::::::|||||||||                |||   |||   |||
Db    808 GAATATGTAAAAAAAATAGTGGGTGAGAGCTATGCAAAATCAACTAAAAAGCGACATACA 867

Qy    142 GlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGln 161
                      :::   :::|||:::     ::::::          |||
Db    868 ATTACTGTAGCTCTAGTTAACGAGTTGAACAACATTAAG------AACGAGTATTTGAAT 921

Qy    162 GluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyr 181
          ::::::   :::      :::      :::   |||:::   ::::::|||
Db    922 AAAATAGTTGAATCAACCTCAGAAAGCCAACTACAGATACTGATGATGGAGAGTCGATCA 981

Qy    182 SerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPhe 201
                :::|||
Db    982 AAAGTAGATGAA------------------------------------------------ 993

Qy    202 AlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArg 221
                      :::|||   ||||||   |||:::   :::
Db    994 ------------GCTGTGTCTAAGTTTGAAAAGGACTCATCTTCT------------TCG 1029

Qy    222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgPro------ 239
             :::   |||   |||                              :::|||
Db   1030 TCAAGTTCAGACTCTTCC---------------------------ACTAAACCGGAAGCT 1062

Qy    240 AlaGluTyrSerLysAspAsnLysProTyrLysPro------------------------ 251
          ::::::   :::|||   |||||||||   :::|||
Db   1063 TCAGATACAGCGAAGCCAAACAAGCCGACAGAACCAGGAGAAAAGGTAGCAGAAGCTAAG 1122

Qy    252 ------ValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMet 269
                |||         |||          :::      :::|||
Db   1123 AAGAAGGTTGAAGAAGCTGAGAAAAAAGCCAAGGATCAAAAAGAAGAAGATCGTCGT--- 1179

Qy    270 ThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArg 289
                   :::|||      |||   :::   |||                  |||
Db   1180 ------AACTACCCAACCATTACTTACAAAACGCTTGAACTTGAAATTGCTGAGTCCGAT 1233

Qy    290 IleGlu---------------------------------AsnGluAsnAsnProArgIle 298
          :::|||                                    |||   ||||||
Db   1234 GTGGAAGTTAAAAAAGCGGAGCTTGAACTAGTAAAAGTGAAAGCTAACGAACCTCGAGAC 1293

Qy    299 GluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
          |||   :::   |||||||||            :::               :::||||||
Db   1294 GAGCAAAAA---ATTAAGCAAGCAGAAGCGGAAGTTGAGAGTAAACAAGCTGAGGCTACA 1350

Qy    319 ArgIleLys---------------------TyrAlaSerLysTyrAlaGlnSerAlaAsn 331
          |||:::|||                        |||   :::   |||:::
Db   1351 AGGTTAAAAAAAATCAAGACAGATCGTGAAGAAGCAGAAGAAGAAGCTAAACGAAGAGCA 1410

Qy    332 TyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArg 351
                |||         |||      :::|||   |||:::         |||   |||
Db   1411 GATGCTAAAGAGCAA---GGTAAACCAAAGGGGCGGGCAAAACGAGGAGTTCCTGGAGAG 1467

Qy    352 LysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaVal 371
                      ::::::            |||         |||:::      |||:::|||
Db   1468 CTAGCAACACCTGATAAAAAAGAAAATGAT------GCGAAGTCTTCAGATTCTAGCGTA 1521

Qy    372 TyrGlyAspValLeuSerSerLeuGlu---LysAlaTyrLysGluGlyAlaLysAlaAsn 390
                :::   |||   |||            |||      |||:::   |||:::|||
Db   1522 GGTGAAGAAACTCTTCCAAGCCCATCCCTGAAACCAGAAAAAAAGGTAGCAGAAGCTGAG 1581

Qy    391 ArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAla 410
          ::::::::                 :::                  |||      |||
Db   1582 AAGAAGGTTGAAGAAGCTAAGAAAAAAGCCGAGGATCAAAAAGAAGAAGATCGCCGTAAC 1641

Qy    411 GlnPheAlaAsnAlaLeuAlaThr------------------------AsnProAspAla 422
                   |||         |||                           :::   :::
Db   1642 TACCCAACCAATACTTACAAAACGCTTGAACTTGAAATTGCTGAGTCCGATGTGGAAGTT 1701

Qy    423 HisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAsp 442
                         |||:::   :::   ::::::   |||:::         |||      :::
Db   1702 AAAAAAGCGGAGCTTGAACTAGTAAAAGAGGAAGCTAAGGAA------CCTCGAAACGAG 1755

Qy    443 ArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeu 462
             ||||||      |||      ::::::      ::::::         |||   :::|||
Db   1756 GAAAAAGTTAAGCAAGCAAAAGCGGAAGTTGAGAGTAAAAAAGCTGAGGCTACAAGGTTA 1815

Qy    463 ProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAla 482
                                     :::|||   |||   |||   ||||||
Db   1816 --------------------------GAAAAAATCAAGACAGATCGTAAAAAAGCAGAA 1848

Qy    483 AspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLys 502
          :::                                   |||      |||      :::
Db   1849 GAA-----------------------------GAAGCTAAACGAAAAGCAGCAGAAGAA 1878

Qy    503 SerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer 522
                   |||||||||   |||:::         :::      ||||||   :::
Db   1879 GATAAAGTTAAAGAAAAAACCAGCTGAACAACCACAACCAGCGCCGGCTCCAAAAGCAGAA 1938

Qy    523 LysSerValIleAla---------------AlaAlaArgAlaIleGlnAlaAspAlaMet 537
          |||         |||                  |||         :::|||:::
Db   1939 AAACCAGCTCCAGCTCCAAAACCAGAGAATCCAGCTGAACAACCAAAAGCAGAAAAACCA 1998

Qy    538 AlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhe---------PheAlaGlyLeu 554
          |||:::      |||   |||         |||                  :::   |||
Db   1999 GCTGATCAACAAGCTGAAGAAGACTATGCTCGTAGATCAGAAGAAGAATATAATCGCTTG 2058

Qy    555 ArgGluMetTyrProGlyArg------------AlaLeuProSerAspAlaAsnPheThr 570
             :::      |||   :::                  |||   ||||||
Db   2059 ACTCAACAGCAACCGCCAAAAACTGAAAAACCAGCACAACCATCTACTCCAAAA------ 2112

Qy    571 MetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsn 590
                                        |||::::::   ::::::|||   ||||||
Db   2113 -----------------------ACAGGCTGGAAACAAGAAAACGGTATGTGGTACTTC 2148

Qy    591 TyrHisThrThrGlyLys---------GlyValLeuGluLysGlnAspProLysSerAsp 607
          |||:::|||   |||            |||   |||:::
Db   2149 TACAATACTGATGGTTCAATGGCGACAGGATGGCTCCAA--------------------- 2187

Qy    608 GluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyr 627
                                                         |||   |||   :::
Db   2188 -------------------------------------------AACAATGGCTCATGG 2202

Qy    628 ---------AlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAsp------ 642
                            ||||||   :::         :::|||   ||||||
Db   2203 TACTACCTCAACAGCAATGGCGCTATGGCGACAGGATGGCTCCAAAACAATGGTTCATGG 2262

Qy    643 --------------------------IleThrGly------GlyAsnSerGlySerPro 651
                                    ||||||         |||:::||||||
Db   2263 TACTATCTAAACGCTAATGGTTCAATGGCAACAGGATGGCTCCAAAACAATGGTTCATGG 2322

Qy    652 ValPhe---AspLysAsnGlyArgLeuIle------GlyLeuAlaPheAspGlyAsnTrp 668
             :::   :::   ||||||   :::            |||      ::::::|||:::|||
Db   2323 TACTACCTAAACGCTAATGGTTCAATGGCGACAGGATGGCTCCAATACAATGGCTCATGG 2382

Qy    669 GluAlaMetSerGlyAsp-IleGluPheGluProAspLeuGlnArgThrIleSerValAs 688
                ::::::   :::   :::::::::            |||         |||::::::
Db   2383 TACTACCTAAACGCTAATGGTTCAATGGCGACAGGATGGCTCCAATACAATGGCTCATGG 2442

Qy    688 pIleArgTyrValLeuPheMetIleAspLysTrpGlyGln 701
                   |||   :::|||      |||   |||
Db   2443 TACTACCTAAACGCTAATGGTGATA------TGGCGACAG 2476

RESULT 11
US-09-134-078-14
; Sequence 14, Application US/09134078
; Patent No. 6368844
;  GENERAL INFORMATION:
;    APPLICANT: Bylina, Edward J.
;    TITLE OF INVENTION:  GLYCOSIDASE ENZYMES
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  Gray Cary Ware & Freidenrich LLP
;      STREET:  4365 Executive Drive, Suite 1600
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/134,078
;      FILING DATE:  13-AUG-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/949,026
;      FILING DATE:  10-OCT-1997
;      APPLICATION NUMBER:  60/056,916
;      FILING DATE:  06-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09010/024002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  858/677-1456
;      TELEFAX:  858/677-1465
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...2316
US-09-134-078-14

Alignment Scores:
Pred. No.:               0.0475        Length:        2319
Score:                   110.00        Matches:       155
Percent Similarity:      33.12%        Conservative:  105
Best Local Similarity:   19.75%        Mismatches:    268
Query Match:             2.96%         Indels:        258
DB:                      4             Gaps:          42

US-10-008-355-2 (1-712) x US-09-134-078-14 (1-2319)

Qy     74 ThrGly---IleThrValSerAspGlnGlyLeuIlePheThrAsnHisHisCysGlyTyr 92
          ||||||   :::           |||:::|||     |||::::::      |||||||||
Db    174 ACAGGCAAGGTCGAACAAGGTGATCGAGGC------TTTTCTGACCAATCCTGTGGATAC 227

Qy     93 GlyAlaIleGlnSer---GlnSer-ThrValAspHisAspTyrLeuArgAspGlyPheVa 111
                         ::::::   |||    |||||||||
Db    228 GAAAAAGAAAGAACTCTTCAAGGTTACTGTTGAC------------------------- 261

Qy    111 lSerArgThrMetGlyGluGluLeuProIleProGlyLeuSer---------------- 125
                          |||:::|||:::|||:::      :::
Db    262 ------------GGAAAAGAGATTCCCGTCTCAAGAGTGGAAAAGGCCGATCCCACGGA 308

Qy    126 ---------ValLysTyrLeuArgLysIleValLysValThrAspLysValGluGlyGl 142
                          |||:::|||   ::::::      :::   |||   |||
Db    309 CATAGACGTGACGAACTACGTGAGAATCGTCCTTTCTGAATCCCTGAAAGAAGAAGACCT 368

Qy    142 nLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGl 162
             |||   :::       :::|||      :::   |||:::   :::         ||
Db    369 CAGAAAAGACGTGGAACTGATCATAGAAGGTTACAAACCGGCAAGAGTCATCATGATGGA 428

Qy    162 uLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyrSe 182
           |                                        |||::::::   :::|||
Db    429 G---------------------------------------ATCCTGGACGACTACTATTA 449

Qy    182 rAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPheAl 202
           :::   |||         ||||||         |||   :::      |||:::::
Db    450 CGATGGAGAG---CTCGGAGCCGTATATTCTCCAGAGAAGACGATATTCAGAGTCTGGTC 506

Qy    202 aProProSerSer---------------ValGlyLysPheGlyGlyAspThrAsp----- 215
          :|||   |||                   :::   |||   |||   ||||||:::
Db    507 CCCCGTTTCTAAGTGGGTAAAGGTGCTTCTCTTCAAAAACGGAGAAGACACAGAACCGTA 566

Qy    216 ---------------------------AsnTrpMetTrpProArgHisThr---GlyAs 225
                                      |||   :::|||           |||||
Db    567 CCAGGTTGTGAACATGGAATACAAGGGAAACGGGGTCTGGGAAGCGGTTGTTGAAGGCGA 626

Qy    225 pPhe---SerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLy 244
          |         ||||||   :::|||               :::        |||   ||
Db    627 TCTCGACGGAGTGTTCTACCTCTAT---------------CAGCTGGAAAACTACGGAAA 671

Qy    244 sAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAl 264
          |                  |||   |||      ||||||                 |||
Db    672 GATCAGAACAACCGTCGATCCT---TATTCGAAAGCGGTT------------TACGCAAA 716

Qy    264 aAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTr 284
           :::      :::      :::         :::                       ||
Db    717 CAACCAAGAGAGCGCCGTTGTGAATCTTGCCAGGACAAACCCAGAA---------GGATG 767

Qy    284 pGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGlyIleLy 304
          |                  |||||||:::      |||:::||||||      |||   ::
Db    768 G-----------------GAAAACGACAGGGGACCGAAAATCGAA------GGATACGA 803

Qy    304 sGlnGlyIleTrpLysGluAlaMetSerAlaAsp-----------GlnAlaThrArgIl 320
          :      |||      |||         ||||||           :::   :::   ::
Db    804 AGACGCGATAATCTATGAAATACACATAGCGGACATCACAGGACTCGAAAACTCCGGGGT 863

Qy    320 eLysTyrAlaSerLysTyr---AlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMe 339
          :|||            |||         :::                  |||      |||::
Db    864 AAAAAACAAAGGCCTCTATCTCGGGCTCACCGAAGAAAACACGAAAGGACCGGGCGGTGT 923

Qy    339 tAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPh 359
          :      ||||||:::   |||         :::|||                    ||
Db    924 GACAACAGGCCTTTCGCACCTTGTGGAACTCGGTGTTACACACGTTCATATACTTCCTTT 983

Qy    359 eAlaAspTrp--------------IleArgLysAsnGlyLysSerAlaValTyr----- 372
          |   |||:::               :::   |||:::   :::         :::
Db    984 CTTTGATTTCTACACAGGCGACGAACTCGATAAAGATTT-CGAGAAGTACTACAACTGGG 1042

Qy    373 ------------------------------GlyAspValLeuSerSerLeuGluLysAl 382
                                        |||   :::|||:::      :::|||
Db   1043 GTTACGATCCTTACCTGTTCATGGTTCCGGAGGGCAGATACTCAACCGATCCCAAAAACC 1102

Qy    382 aTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGl 402
              |||      ::::::   |||                  |||:::|||   :::
Db   1103 CACACACGAGAATCAGAGAAGTCAAAGAAATGGTCAAAGCCCTTCACAAACACGGTATAG 1162

Qy    402 y---------GlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAs 419
                      |||      |||            ::::::            |||
Db   1163 GTGTGATTATGGACATGGTGTTCCCTCACACCTACGGTATAGGCGAACTCTCTGCGTTCG 1222

Qy    419 nProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLys---------------- 433
            |||               :::|||::::::      |||:::
Db   1223 ATCAGACGGTGCCGTACTACTTCTACAGAATCGACAAGACAGGTGCCTATTTGAACGAAA 1282

Qy    434 ---------------------------------TyrLysAspTyrLeuProSerLeuAs 442
                                           :::   :::   :::   |||
Db   1283 GCGGATGTGGTAACGTCATCGCAAGCGAAAGACCCATGATGAGAAAATTCATAGTCGATA 1342

Qy    442 pArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAla-AspLysL 462
           |||   :::|||                  :::      |||||||||      |||:::
Db   1343 CCGTCACCTACTGGGTAAA------GGAGTATCACATAGACGGATTCAGGTTCGATCAGA 1396

Qy    462 euProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyr- 481
          ::                     :::||||||||||||         :::   ::::::
Db   1397 TG--------------GGTCTCATCGACAAAAAGACAATGCTCGAAGTCGAAAGAGCTC 1441

Qy    482 ------------------------------------------------------AlaA 483
```

```
                                                                            |||
Db   1442 TTCATAAAATCGATCCAACTATCATTCTCTACGGCGAACCGTGGGGTGGATGGGGAGCAC 1501

Qy    483 spPheValPheAspLysSer------------------ValValProTyrSerAspLysP 497
                            |||   ||||||                  |||      :::::|||:::|
Db   1502 CGATCAGGTTTGGAAAGAGCGATGTCGCCGGCACACACGTGGCAGCTTTCAACGATGAGT 1561

Qy    497 heHisAlaMetLeuLys---------------------------------------S 503
                    ||            :::::
Db   1562 TCAGAGACGCAATAAGGGGTTCCGTGTTCAACCCGAGCGTCAAGGGATTCGTCATGGGAG 1621

Qy    503 erMetAspLysGlu---------------------------------------------L 508
                            |||||                                              |
Db   1622 GATACGGAAAGGAAACCAAGATCAAAAGGGGTGTTGTTGGAAGCATAAACTACGACGGAA 1681

Qy    508 ysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer---------------- 522
                    ||      |||:::         ||||||         ::::::
Db   1682 AACTCATCAAAAGTTTCGCCCTTGATCCAGAAGAAACTATAAACTACGCAGCGTGTCACG 1741

Qy    523 --------------------LysSerValIleAlaAla------------AlaArgAlaI 532
                                        |||:::   :::||||||               :::
Db   1742 ACAACCACACACTGTGGGACAAGAACTACCTTGCCGCCAAAGCTGATAAGAAAAAGGAAT 1801

Qy    532 leGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheA 552
                            :::   :::   ||||||      :::                |||
Db   1802 GGACCGAAGAAGAACTGAAAAACGCCCAGAAACTGGCTGGTGCGATACTTCTCACTTCTC 1861

Qy    552 laGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetA 572
                      |||:::      ::::::   |||:::                  ||||||
Db   1862 AAGGTGTTCCTTTCCTCCACGGAGGGCAGGACTTCTGCAGGACGACGAATTTCAACGACA 1921

Qy    572 rgMetSerTyrGly---------SerIleLysGlyTyrGluProGlnAspGlyAlaTrp- 588
                         ||||||            ||||||   |||::::::   :::
Db   1922 AC---TCCTACAACGCCCCTATCTCGATAAACGGCTTCGATTACGAAAGAAAACTTCAGT 1978

Qy    589 ----------TyrAsnTyrHisThrThrGlyLysGlyVal------LeuGluLysGlnA 603
                               :::|||||||||         ||||||:::      |||   |||:::
Db   1979 TCATAGACGTGTTCAATTACCAC---------AAGGGTCTCATAAAACTCAGAAAAGAAC 2029

Qy    603 spProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysA 623
                      |||                                          ||||||   ||||
Db   2030 ACCCT-------------------------------------GCTTTCAGGCTGAAAA 2050

Qy    623 snTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn------- 640
                    ||            ||||||      :::   |||:::   ||||||
Db   2051 AC-----------GCTGAAGAGATCAAAAAACACCTGGAATTTCTCCCGGGCGGGAGAA 2098

Qy    641 -------------------AsnAspIleThrGlyGlyAsnSerGlySerPro---ValP 653
                                           |||      ||||||:::               |||
Db   2099 GAATAGTTGCGTTCATGCTTAAAGACCACGCAGGTGGTGATCCCTGGAAAGACATCGTGG 2158

Qy    653 heAspLysAsnGlyArgLeuIleGlyLeuAlaPhe---------AspGlyAsnTrp---- 668
                            ||||||   |||            :::         :::|||   |||
Db   2159 TGATTTACAATGGAAACTTAGAGAAGACAACATACAAACTGCCAGAAGGAAAATGGAATG 2218

Qy    669 -------------------------------------GluAlaMetSerGlyAspIleG 676
                                                  |||   :::    |||  ||||
Db   2219 TGGTTGTGAACAGCCAGAAAGCCGGAACAGAAGTGATAGAAACCGTCGAAGGAACAATAG 2278

Qy    676 luPheGluPro 679
                    ||   :::|||
Db   2279 AACTCGATCCG 2289
```

RESULT 12
US-09-134-001C-195
; Sequence 195, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 195
; LENGTH: 2667
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-195

```
Alignment Scores:
Pred. No.:               0.059          Length:        2667
Score:                   110.00         Matches:       112
Percent Similarity:      34.72%         Conservative:  89
Best Local Similarity:   19.34%         Mismatches:    178
Query Match:             2.96%          Indels:        201
DB:                      4              Gaps:          28

US-10-008-355-2 (1-712) x US-09-134-001C-195 (1-2667)

Qy    183 AsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPheAla 202
          :::|||   ||||||            |||      |||:::         :::
Db    100 AGTAATGGTTACTTTAAACCATCTGAAGATAAATCTAAAGAGGCATATACAATTGTAATT 159

Qy    203 ProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHis 222
          ||||||                                             |||
Db    160 CCGCCC---------------------------------------------CCTAATGTA 174

Qy    223 ThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyr 242
          ||||||         :::         :::
Db    175 ACAGGTAAATTACATTTAGGTCATGCATGG------------------------------ 204

Qy    243 SerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyr 262
                   :::         :::   :::                  |||||||||||
Db    205 ---------GATACTACTTTACAAGATATAATAACTAGAATGAAAAGAATGCAAGGATAC 255

Qy    263 LysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThr 282
                                 |||:::      ||||||
Db    256 GAT------------------ACTTTATATCTGCCTGGTATG------------------ 279

Qy    283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                         |||                        :::::|||   :::
Db    280 ---------------GATCATGCTGGTATAGCAACTCAAGCAAAGGTTGAAGCAAAACTT 324

Qy    303 IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLysTyr 322
          :::||||||||||     :::      :::      :::
Db    325 AATGAACAAGGTATATCTAGACACGACTTAGGAAGAGAA--------------------- 363

Qy    323 AlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsn--------------------- 335
                |||:::   |||   |||      ||||||
Db    364 ------AAATTTTTACAGCAAGCATGGGATTGGAAAGAAGAGTATGCAACATTTATTAGA 417

Qy    336 ------------SerIleGlyMetAsnArgGlyLeuAlaArg------------LeuAsp 347
                         :::|||:::            :::|||            ||||||
Db    418 CAACAATGGGCTAAACTTGGCTTAGGGTTAGATTATAGTAGAGAACGCTTTACGTTAGAT 477

Qy    348 ValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLys------ 365
                |||      :::|||      :::   |||   |||         |||
Db    478 ---GATGGTTTAAGTAAAGCTGTAAGAAAAGTATTTGTTGATTATATAATAAAGGTATT 534

Qy    366 ------AsnGlyLysSerAlaValTyrGlyAsp---------------ValLeuSerSer 378
                   |||:::      :::      |||                  ||||||
Db    535 ATTTATCGTGGTGAAAGAATTATTAACTGGGACCCTATAGCTAGAACAGCTTTATCAGAT 594

Qy    379 LeuGluLysAlaTyrLys------GluGlyAlaLysAlaAsnArgGluMetThrTyr--- 395
          :::|||      ::::::      :::||||||      :::   :::      |||
Db    595 ATTGAAGTCATTCATGAAGATGTCCAAGGTGCTTTTTATCATTTTAAATATCCTTATGGT 654

Qy    396 ---------------------------------LeuSerGluThrLeuPheGlyGlyThr 404
                                                   ||||||:::  |||   |||
Db    655 GATGGGAATGGATATATAGAAATAGCTACTACACGTCCTGAGACTATGTTAGGTGACACT 714
```

```
Qy    405 GluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla 424
                                        |||:::      ||||||:::
Db    715 ------------------------------GCTATTGTTGTAAATCCAAAT--------- 735

Qy    425 GlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLys 444
                            |||:::::||||||||     :::
Db    736 ------------------GACGAAAGGTATAAAGATGTGATTGGTAAA------------ 765

Qy    445 ValLeuProAlaMetLeuAspIleValArgArgArgIlePro------AlaAspLysLeu 462
                      :::|||   ||||||   |||   :::|||      ||||||:::
Db    766 ------ACAGTTATATTACCTATTGTAGGAAGAGAGTTACCTATTTTAGCTGATGAATAC 819

Qy    463 ProAspIle--------------------------------------------------- 465
             ||||||
Db    820 GTTGATATTGAATTTGGTTCAGGAGCAATGAAAGTTACTCCGGCACATGATCCTAATGAC 879

Qy    466 -----------------------PheLysAsnValIle------AspLysLysPheLys 475
                                     :::|||:::|||      |||:::      |||
Db    880 TTTGAAATTGGTCAAAGACATCAATTAGAAAATATTATAGTCATGGATGAATATGGTAAG 939

Qy    476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
             :::   |||                                          :::|||
Db    940 ATGAACGATAAA------------------------------------------GCTGAT 957

Qy    496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIle---Glu 514
          |||:::      |||   :::         :::   :::      |||   :::   |||
Db    958 AAATACAAAGGGATGGATAGATTTGATTGTAGAAATCAGCTTGTTAAAGACTTAAAAGAA 1017

Qy    515 LysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAla 534
          :::|||      ::::::::    :::         :::         :::   ::::::
Db   1018 CAAGATCTTGTTATAAAAATTGAAGAACATACACATTCAGTAGGACATTCTGAACGTTCT 1077

Qy    535 AspAlaMetAlaAsnAlaTyr--------------------------------------- 541
             |||:::         |||
Db   1078 GGTGCAATAGTAGAACCTTATTTGTCTACACAATGGTTTGTAAAAATGAAACCTTTGGCA 1137

Qy    542 ------AlaIleGlu-----------LysGlyLysArgLeuPhePheAlaGlyLeuArg 555
                |||:::::             |||   :::      ||||||  |||
Db   1138 CAAAGAGCTTTGGATAATCAAAATACAAAAGATAGAATTGACTTTTTCCCAGGTAGATTT 1197

Qy    556 GluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMet-ArgMetSerTy 575
          |||            |||   :::            :::::|||:::  |||::::::||
Db   1198 GAAAATACATTCAATCGCTGGATGGAAGAAATTAGAGATTGGACGATTTCGCGTCAACTA 1257

Qy    575 rGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGl 595
          ||||                             |||||||||    :::      |||||
Db   1258 TGGTGGGGCCACCAAATACCT------------GCTTGGTAT---CATAAAGATACAGG 1301

Qy    595 yLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln---GluAs 614
          |:::      :::   ::::::   |||:::      |||      :::|||   |||::
Db   1302 TGAAGTCTTTGTTGGTGAAGAAGCACCTGAAGACATCGAGAATTGGATACAAGATGAGGA 1361

Qy    614 nIleLeuAspLeuPheArgThrLys-AsnTyrGlyArgTyrAlaGlu---AsnGlyGlnL 633
          :::||||||    :::   :::    ::::::||||||:::         :::|||   :
Db   1362 TGTATTAGATACATGGTTCTCAAGTGCACTTTGGCCGTTTTCAACATTAGGATGGCCTGA 1421

Qy    633 euHisIleAlaPheLeuSerAsnAsnAspIle--------------------------T 644
          ::   :::      :::   |||      |||
Db   1422 TACAAATGCTGATGATTTAAACGTTATTATCCGACAAATGCATTAGTTACTGGTTATGA 1481

Qy    644 hrGlyGlyAsnSerGlySerProVal---PheAspLysAsnGlyArgLeuIle 660
                  :::||||||   ||||||   |||   |||:::      |||:::
Db   1482 TATCATTTTCTTCTGGGTAGCCCGTATGATTTTCCAAGGATTAGAATTTACTG 1534
```

```
RESULT 13
US-08-446-137B-1
; Sequence 1, Application US/08446137B
; Patent No. 6162903
;  GENERAL INFORMATION:
;    APPLICANT:  Trowern, Angus R.
;    APPLICANT:  Atkinson, Anthony
;    APPLICANT:  Murphy, Jonathan P.
;    APPLICANT:  Laurence, Oliver S.
;    APPLICANT:  Duggleby, Clive J.
;    TITLE OF INVENTION:  IMMUNOGLOBULIN BINDING PROTEINS DERIVED
;    TITLE OF INVENTION:  FROM L PROTEIN AND THEIR USES
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,137B
;      FILING DATE:  22-MAY-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  100084.406
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3279 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Peptostreptococcus asaccharolyticus
;      STRAIN:  1018
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  103..3186
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 280
;      OTHER INFORMATION:  /product= "mature protein L"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..3186
;      OTHER INFORMATION:  /codon_start= 103
;      OTHER INFORMATION:  /product= "immature protein L"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  280..3183
;      OTHER INFORMATION:  /codon_start= 280
;      OTHER INFORMATION:  /product= "mature protein L"
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  208..279
US-08-446-137B-1

Alignment Scores:
Pred. No.:                0.0812        Length:         3279
Score:                    110.00        Matches:        118
Percent Similarity:       31.59%        Conservative:   63
Best Local Similarity:    20.59%        Mismatches:     206
Query Match:              2.96%         Indels:         186
DB:                       4             Gaps:           29

US-10-008-355-2 (1-712) x US-08-446-137B-1 (1-3279)
```

```
Qy     29 TrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPheThr 48
          |||||||||          :::|||   :::    ::::::|||    |||    |||
Db   1861 TGGTTATTAAAGAATGCTAAAGAAGAAGCAATCAAAGAATTAAAAGAAGCAGGAATCACT 1920

Qy     49 LeuProLeuAspSerLeuTyrSerPheAspLysProSerIleAlaAsnAlaValValIle 68
                            ||||||   |||         :::   |||      |||
Db   1921 ---------TCTGATTTATAC---TTCAGCTTAATCAATAAAGCAAAAACAGTT------ 1962

Qy     69 PheGlyGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
                         |||:::                     ::::::
Db   1963 ---------------GAAGGCGTAGAAGCATTAAAGAACGAAATCTTA------AAAGCA 2001

Qy     89 HisCysGlyTyrGlyAlaIleGlnSerGlnSerThrValAspHisAspTyrLeuArgAsp 108
          |||   |||                       :::   |||   :::      |||:::|||
Db   2002 CACGCTGGA-----------------GAAGAAACACCAGAA---------TTAAAAGAT 2034

Qy    109 GlyPheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSerValLysTyr 128
          |||:::   :::               |||                  |||
Db   2035 GGATATGCAACATATGAAGAAGCAGAAGCAGCAGCTAAAGAAGCTTTGAAAAATGATGAT 2094

Qy    129 LeuArgLysIleValLysValThrAspLysValGluGlyGln-----------LeuLys 144
          :::         ::::::             :::|||:::              ||||||
Db   2095 GTTAACAACGCATACGAAATAGTTCAAGGTGCAGACGGAAGATACTACTATGTATTAAAG 2154

Qy    145 ---GlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGln------ 161
             :::  ||||||   |||                   ||||||   :::
Db   2155 ATTGAAGTTGCAGACGAAGAAGAACCAGGTGAAGACACTCCAGAAGTTCAAGAAGGTTAC 2214

Qy    162 -----------------GluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCys 175
                           |||   |||   ||||||          |||:::::::::
Db   2215 GCAACTTACGAAGAAGCAGAAGCAGCAGCTAAAGAAGCATTAAAAGAAGATAAAGTT--- 2271

Qy    176 IleValGluProPheTyrSerAsnAsnGluTyrPheLeuIle------------------ 189
                               ||||||   |||   ::::::
Db   2272 --------------------AACAATGCATACGAAGTAGTTCAAGGTGCAGACGGAAGA 2310

Qy    190 ValTyrAspValPheLys---AspValArgMetValPheAlaProProSerSerValGly 208
             |||   |||   |||   :::   :::                |||          |||
Db   2311 TACTACTATGTATTAAAAATCGAAGATAAAGAAGATGAACAACCAGGTGAAGAACCAGGC 2370

Qy    209 LysPheGlyGlyAspThr---AspAsnTrpMetTrpProArgHisThrGlyAspPheSer 227
          :::      |||   |||   |||   |||:::                |||
Db   2371 GAAAACCCAGGAATCACAATTGATGAATGGTTATTAAAGAATGCTAAAGAAGAC------ 2424

Qy    228 ValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLys 247
                         |||      :::   :::   |||      |||   |||
Db   2425 ---------------GCAATCAAAGAATTAAAAGAAGCAGGAATCAGTTCTGAC------ 2463

Qy    248 ProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyr 267
                      :::||||||   |||:::
Db   2464 -----------ATATACTTTGATGCAATC----------------------------- 2481

Qy    268 AlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGlu 287
                                             ::::::   |||   |||||||||
Db   2482 -----------------------------AACAAAGCAAAAACAGTAGAAGGCGTAGAA 2511

Qy    288 AspArgIleGluAsnGlu-----------------------------AsnAsnProArg 297
             ::::::||||||                                ||||||
Db   2512 ---GCGTTAAAGAACGAAATCTTAAAAGCACACGCTGAAAAACCAGGCGAAAACCCAGGA 2568

Qy    298 IleGluVal---------------ArgGlyIleLysGlnGlyIleTrpLysGlu------ 310
          |||   :::               :::      |||:::         ||||||
Db   2569 ATCACAATTGATGAATGGTTATTAAAGAATGCTAAAGAAGCTGCAATCAAAGAATTAAAA 2628

Qy    311 ------AlaMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGln 328
                ::::::|||:::             :::         :::|||       :::
Db   2629 GAAGCAGGAATCACTGCTGAATATCTATTCAACTTAATCAACAAAGCAAAAACAGTAGAA 2688

Qy    329 SerAlaAsnTyrTrpLysAsnSerIle-----------------------GlyMetAsn 340
                            ||||||   |||                       |||   |||
Db   2689 GGCGTAGAATCATTAAAGAACGAAATCTTAAAAGCACACGCTGAAAAACCAGGCGAAAAC 2748

Qy    341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360
             |||:::
Db   2749 CCAGGAATC----------------------------------------ACAATTGAT 2766

Qy    361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380
          :::|||:::   ||||||   |||                                |||
Db   2767 GAATGGTTATTAAAGAACGCTAAA--------------------------------GAA 2793

Qy    381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400
             |||   ||||||               :::|||            |||:::
Db   2794 GATGCAATTAAAGAA------------TTAAAAGAAGCAGGAATTACTTCTGACATATAC 2841

Qy    401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeu------------ 416
          |||               :::      :::         ||||||
Db   2842 TTTGATGCTATCAACAAAGCAAAAACTATTGAAGGCGTAGAAGCATTAAAGAATGAAATC 2901

Qy    417 ------AlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
                |||         |||         |||         |||   |||
Db   2902 TTAAAGGCTCATAAAAAAGATGAAGAACCAGGTAAAAAACCAGGTGAAGACAAAAAACCA 2961

Qy    435 LysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArg 454
          :::|||         |||      |||:::|||
Db   2962 GAAGATAAAAAACCAGGTGAAGATAAAAAA----------------------------- 2991

Qy    455 ArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPhe 474
                   |||   ||||||   |||         |||         |||||||||
Db   2992 ---------CCAGAAGACAAAAAACCTGGTGAAGATAAAAAACCAGAAGACAAAAAA--- 3039

Qy    475 LysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSer 494
             |||   |||   |||
Db   3040 CCAGGTAAAACAGATAAA------------------------------------------ 3057

Qy    495 AspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGlu 514
                                       |||   :::|||:::|||   ||||||
Db   3058 --------------------GATTCTCCAAATAAGAAGAAAAAAGCTAAATTA------ 3090

Qy    515 LysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAla 534
                |||                     ::::::   |||   |||         :::
Db   3091 ------CCAAAAGCTGGTAGCGAAGCTGAAATCTTAACATTAGCAGCAGCAGCTTTATCA 3144

Qy    535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLys 547
             |||               :::::::::|||   |||
Db   3145 ACAGCAGCAGGTGCTTACGTTTCACTTAAAAAACCTAAA 3183
```

RESULT 14
US-08-961-527-139/c
; Sequence 139, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders

```
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  139:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15363 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-139

Alignment Scores:
Pred. No.:               1.14          Length:        15363
Score:                   109.00        Matches:       124
Percent Similarity:      29.03%        Conservative:  83
Best Local Similarity:   17.39%        Mismatches:    222
Query Match:             2.93%         Indels:        284
DB:                      4             Gaps:          30

US-10-008-355-2 (1-712) x US-08-961-527-139 (1-15363)

Qy      210 PheGlyGlyAsp-----------------------ThrAspAsnTrpMetTrpProArg 221
             :::|||  |||                        |||   :::|||   :::   |||
Db    10583 TATGGAAATGATCGTGTGGGTTGGGGAAAACAGCAACCGCTTCATGGACTTTTGACCGT 10524

Qy      222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAsp------AsnArgPro 239
                     ::::::         ::::::  |||   |||              |||
Db    10523 GACAACGCTGGCTATGCTGGACAGTTTATCTGGACAGGTACGGACTATATTGGTGAACCT 10464

Qy      240 AlaGluTyrSerLysAspAsnLys---ProTyrLysProValTyrPheAlaAlaValSer 258
                  :::             |||:::  |||   |||      ||||||     |||
Db    10463 ACACCATGGCACAACCAAAATCAAACTCCTGTTAAGAGCTCTTACTTTGGTATCGTAGAT 10404

Qy      259 MetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAsp 278
                  |||          |||:::   :::                     |||
Db    10403 ACAGCCGGCATTCCAAAACATGACTTCTATCTCTACCAAAGCCAATGGGTTTCTGTTAAG 10344

Qy      279 Arg---------------TyrLeuThrSerTrpGlyValGluAspArg------------ 289
            :::                      |||     |||    |||::::::
Db    10343 AAGAAACCGATGGTACACCTTCTTCCTCACTGGAACTGGGAAAACAAAGAATTAGCATCC 10284

Qy      290 ------------------------------------------IleGluAsnGluAsnAsnPro 296
                                                       |||   ::::::
Db    10283 AAAGTAGCTGACTCAGAAGGTAAGATTCCAGTTCGTGCTTATTCGAATGCTTCTAGTGTA 10224

Qy      297 ArgIleGluValArgGlyIleLysGlnGlyIle--------------------------- 307
               :::   :::   |||         |||:::
Db    10223 GAATTGTTCTTGAATGGAAAATCTCTTGGTCTTAAGACTTTCAATAAAAAACAAACCAGC 10164

Qy      308 -------------------------------------------------TrpLys--- 309
                                                             ||||||
Db    10163 GATGGGCGGACTTACCAAGAAGGTGCAAATGCTAATGAACTTTATCTTGAATGGAAAGTT 10104

Qy      310 --------------------GluAlaMetSerAlaAspGlnAla-------------- 317
                                ||||||::::::   |||::::::
Db    10103 GCCTATCAACCAGGTACCTTGGAAGCAATTGCTCGTGATGAATCTGGCAAGGAAATTGCT 10044

Qy      318 ---ThrArgIleLysTyrAlaSerLys-------------------------------- 325
                  :::|||      |||   |||
Db    10043 CGAGATAAGATTACGACTGCTGGTAAGCCAGCGGCAGTTCGTCTTATTAAGGAAGACCAT 9984

Qy      326 ------------------------------------TyrAlaGlnSerAlaAsnTyrTrp 333
                                                |||   :::      :::
Db     9983 GCGATTGCAGCAGATGGAAAAGACTTGACTTACATCTACTATGAAATTGTTGACAGCCAG 9924

Qy      334 LysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGly--------- 350
               |||   :::                  |||   |||      :::   |||
Db     9923 GGGAATGTGGTTCCAACTGCTAATAATCTGGTTCGCTTCCAATTGCATGGCCAAGGTCAA 9864

Qy      351 --------------------------------ArgLysArgAlaGluGluArgAlaPhe 359
                                            |||:::|||    :::     :::|||
Db     9863 CTGGTCGGTGTAGATAACGGAGAACAAGCCAGCCGTGAACGCTATAAGGCGCAAGCAGAT 9804

Qy      360 AlaAspTrpIleArgLys------AsnGlyLysSer----------------------- 369
                ||||||||||||    .  |||||||||
Db     9803 GGTTCTTGGATTCGTAAAGCATTTAATGGTAAAGGTGTTGCCATTGTCAAATCAACTGAA 9744

Qy      370 --------------------AlaValTyrGlyAspValLeuSerSerLeuGluLysAla 382
                                    :::   |||:::|||    |||     :::
Db     9743 CAAGCAGGGAAATTCACCCTGACTGCCCACTCTGATCTCTTGAAATCGAACCAAGTCACT 9684

Qy      383 TyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGly 402
                        |||   |||         :::|||   |||   |||
Db     9683 GTCTTTACTGGTAAGAAAGAAGGACAAGAGAAGACTGTTTTG----------------- 9642

Qy      403 GlyThrGluVal----------------------------------------------- 406
            ||||||||||||
Db     9641 GGGACAGAAGTGCCAAAAGTACAGACCATTATTGGAGAGGCACCTGAAATGCCTACCACT 9582

Qy      407 ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIle 426
            |||   |||                    :::   |||      |||         :::   :::
Db     9581 GTTCCGTTTGTATACAGTGATGGTAGCCGTGCAGAACGTCCTGTAACCTGGTCTTCAGTA 9522

Qy      427 --------------------------LeuLysSerLeuAspAspLysTyrLysAspTyr 437
                                      :::|||   :::   |||      ::::::
Db     9521 GATGTGAGCAAGCCTGGTATTGTAACGGTGAAAGGTATGGCTGACGGA---CGAGAAGTA 9465

Qy      438 LeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIle 457
                 ::::::                |||   :::   |||   :::|||:::|||
Db     9464 GAAGCTCGTGTAGAAGTGATTGCTCTTAAATCAGAGCTACCAGTTGTGAAACGTATTGCT 9405

Qy      458 ProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAsp 477
            |||         |||     :::   |||:::|||                  |||
Db     9404 CCAAATACTGACTTGAATTCTGTAGACAAATCTGTTTCCTATGTTTTGATTGATGGAAGT 9345

Qy      478 ThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAspLys--- 496
               :::::::|||                ||||||      :::       ||||||
Db     9344 GTTGAAGAGTATGAA---------GTGGACAAGTGGGAGATTGCCGAAGAAGATAAAGCT 9294

Qy      497 -------------------------------------------------------Phe 497

Db     9293 AAGTTAGCAATTCCAGGTTCTCGTATTCAAGCGACCGGTTATTTAGAAGGTCAACCAATT 9234

Qy      498 HisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspPro 517
            ||||||   |||     .     :::            |||   |||:::         |||
Db     9233 CATGCAACCCTTGTGGTAGAAGAAGGCAATCCTGCGGCACCTGCAGTA---------CCA 9183

Qy      518 AlaValGluLeuSerLysSerValIle------------------------------- 526
               |||   :::         :::
Db     9182 ACTGTAACGGTTGGTGGTGAGGCAGTAACAGGTCTTACTAGTCAAAAACCAATGCAATAC 9123

Qy      527 --------------------------------------------------AlaAla 528
                                                              ||||||
Db     9122 CGCACTCTTGCTTATGGAGCTAAGTTGCCAGAAGTCACAGCAAGTGCTAAAAATGCAGCT 9063

Qy      529 AlaArgAlaIleGlnAlaAspAlaMetAlaAsn------------------------- 539
                  :::||||||   |||   ||||||
Db     9062 GTTACAGTTCTTCAAGCAAGCGCA---GCAAACGGCATGCGTGCGAGCATCTTTATTCAG 9006

Qy      540 -----------------------AlaTyrAlaIle---------GluLysGlyLysArg 548
                                   ||||||||||           |||:::      :::
Db     9005 CCTAAAGATGGTGGCCCTCTTCAAACCTATGCAATTCAATTCCTTGAAGAAGCGCCAAAA 8946

Qy      549 LeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAsp------ 566
            :::            |||   ::::::            :::|||    |||
Db     8945 ATTGCTCACTTGAGCTTG---CAAGTGGAAAAAGCTGACAGTCTCAAAGAAGACCAAACT 8889

Qy      567 AlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGly 586
                  ::::::|||      |||                     ||||||||
Db     8888 GTCAAATTGTCGGTTCGAGCTCACTAT-----------------------CAAGATGGA 8853

Qy      587 AlaTrp--------------------TyrAsnTyrHisThrThrGlyLysGlyValLeu 599
                                      :::   |||:::|||:::|||
Db     8852 ACGCAAGCTGTATTACCAGCTGATAAAGTAACCTTCTCTACAAGTGGTGAAGGG------ 8799
```

```
Qy      600 GluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPhe 619
                                   |||  |||:::::::::   :::|||:::|||
Db     8798 -----------------------GAAGTCGCAATTCGTAAAGGAATGCTTGAGTTGCAT 8763

Qy      620 Arg-----------ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIle 635
             :::           |||  |||                   |||:::   :::
Db     8762 AAGCCAGGAGCAGTCACTCTGAACGCTGAATATGAGGGAGCTAAAGACCAAGTTGAACTC 8703

Qy      636 AlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLys 655
                     :::|||   :::                |||      ||||||
Db     8702 ACTATCCAAGCCAATACTGAGAAGAAGATTGCGCAATCCATCCGTCCTGTA--------- 8652

Qy      656 AsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIle 675
                                                     :::      |||:::
Db     8651 ----------------------------------------AATGTAGTGACAGATTTG 8634

Qy      676 GluPheGluProAspLeuGlnArgThrIleSerValAsp 688
                  ||||||   |||      |||::::::|||:::
Db     8633 CATCAGGAACCAAGTCTTCCAGCAACAGTAACAGTTGAG 8595
```

RESULT 15
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```
Alignment Scores:
Pred. No.:               7.14e+03      Length:         4403765
Score:                   109.00        Matches:        136
Percent Similarity:      29.34%        Conservative:   60
Best Local Similarity:   20.36%        Mismatches:     189
Query Match:             2.93%         Indels:         284
DB:                      4             Gaps:           34
```

US-10-008-355-2 (1-712) x US-09-103-840A-2 (1-4403765)

```
Qy       75 GlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAla 94
               |||  ||||||      |||||||||                 |||   |||:::
Db  4015782 GGTGCCACGGTG------CAGGGCCTA-----------CCGCATATCGGGCAT------ 4015747

Qy       95 IleGlnSerGlnSerThrValAspHisAspTyrLeuArgAspGlyPheValSerArgThr 114
                     :::|||   |||      |||   ||||||         ::::::|||
Db  4015746 ------GTCCGCAGCGGGGTCGCCTTCGACATCCTGCGCCGATGGTTGCTCGCGCGT--- 4015696

Qy      115 MetGlyGluGluLeuProIleProGlyLeuSerValLysTyrLeuArgLysIleValLys 134
                                        |||      |||    ::::::|||    :::
Db  4015695 -----------------------GGCTACGACGTCGCGTTTATCCGCAACGTGACCGAC 4015660

Qy      135 ValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArg 154
               :::  ||||||:::
Db  4015659 ATCGAAGACAAGATC--------------------------------------------- 4015645

Qy      155 LysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeu 174
                                    |||||||||            |||
Db  4015644 ----------------------CTGGCCAAGGCCGCCGCGGCGGGC----------- 4015621

Qy      175 CysIleValGluProPheTyr-----------SerAsnAsnGluTyrPheLeuIleVal 190
                         |||::::::              ::::::          |||
Db  4015620 ---------CGGCCGTGGTGGGAGTGGGCGGCTACCCACGAGCGTGCCTTCACCGCGGCC 4015570

Qy      191 TyrAspValPheLysAspValArgMetValPheAlaProProSerSerValGlyLysPhe 210
             ||||||       ||||||           |||||||||:::
Db  4015569 TACGACGCTCTG---GACGTC-----------TTGCCGCCGTCCGCGGAG--------- 4015534

Qy      211 GlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe--- 229
                                      ||||||   ||||||      :::
Db  4015533 --------------------------CCGCGCGCCACCGGGCATATCACCCAGATGATC 4015501

Qy      230 -----------------------------ArgValTyrAlaGlyAlaAspAsnArgPro 239
                                          |||   |||      :::
Db  4015500 GAGATGATCGAGCGCCTGATCCAGGCCGGCCACGCGTATACCGGTGGCGGTGAC------ 4015447

Qy      240 AlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSer--- 258
                                                 |||||||||      :::|||
Db  4015446 ----------------------------------GTTTACTTCGACGTTCTGAGCTAC 4015423

Qy      259 --------------MetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPhe 273
                      :::   |||:::|||   ||||||                   |||
Db  4015422 CCGGAGTACGGCCAGTTATCCGGTCACAAGATCGATGACGTCCATCAGGGCGAAGGTGTG 4015363

Qy      274 ---ProGlySerThrAsp---ArgTyrLeuThrSerTrpGlyValGlu-AspArgIleGl 291
                  |||      |||   |||      |||   |||      |||   |||:::
Db  4015362 GCTGCCGGCAAGCGGGACCAGCGCGACTTCACTTTGTGGAAGGGCGAAAAGCCGGGTG-- 4015305

Qy      291 uAsnGluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAl 311
              |||            |||            |||::::::   ||||||||
Db  4015304 -AACCGTCGTGGCCGACGCCGTGGGGCCGCGGGCGTCCGGGCTGGCATCTGG-------- 4015254

Qy      311 aMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAs 331
                                                                :::|||
Db  4015253 ----------------------------------------------------AATGCTCG 4015246

Qy      331 nTyrTrp-----------------LysAsnSerIleGlyMetAsnArgGlyLeuAlaAr 345
              |||                 :::|||||||||   :::         |||:::
Db  4015245 GCAATGGCTCGCAGCTATCTCGGGCCGGAATTCGATATCCATTGC-GGTGGAATGGATTT 4015187

Qy      345 gLeu--------------AspValIleGlyArgLysArgAlaGluGluArgAlaPheAl 360
              :::              :::   |||   :::   ||||||              |||||
Db  4015186 AGTCTTCCCGCATCATGAGAACGAAATCGCCCAGAGCCGCGCCGCCGGTGATGGGTTCGC 4015127

Qy      360 a-----------------AspTrpIleArgLysAsnGlyLys------SerAlaValTy 372
             |                      |||:::         |||:::      |||
Db  4015126 CCGCTACTGGCTGCACAATGGCTGGGTGACGATGGGCGGGGAGAAGATGAGCAAGTCGCT 4015067

Qy      372 rGlyAspValLeuSer-------------------------SerLeuGluLysAlaTy 383
              |||:::|||||||||                                |||   ||
Db  4015066 GGGCAACGTGTTGTCCATGCCGGCGATGTTGCAGCGGGTGCGGCCGGCTGAGCTGCGTTA 4015007

Qy      383 rLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGl 403
              |         |||:::         |||   |||         |||||||||
Db  4015006 TTACCTGGGCAGCGCGCACTACCGGTCGATGCTCGAGTTCTCCGAGACC---------- 4014958

Qy      403 yThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHi 423

Db  4014958 ------------------------------------------------------------ 4014958

Qy      423 sAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspAr 443
                              ::::::   |||         |||   |||:::
Db  4014957 ---------------GCTATGCAGGATGCGGTGAAAGCCTATGTCGGG----------- 4014925

Qy      443 gLysValLeuProAlaMetLeuAspIleValArgArgArgIle-----------ProAl 459
                     |||         |||         ||||||   |||:::                |||
Db  4014924 -------CTAGAGGACTTCCTGCACCGCGTTCGCACCCGGGTGGGCGCCGTCTGCCCCGG 4014872

Qy      459 aAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLy 479
```

```
             |||       |||
Db 4014871 CGATCCGACCCCA-------------------------------------------- 4014859

Qy     479 sLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAl 499

Db 4014859 --------------------------------------------------------- 4014859

Qy     499 aMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaVa 519
                                       :::||||||:::|||::::::    |||    :::||
Db 4014858 ------------------------CGGTTCGCCGAAGCGCTCGACGACGACCTGTCCGT 4014824

Qy     519 lGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAs 539
            |    ::::::   :::   |||         ||||||
Db 4014823 TCCGATCGCGCTCGCCGAGATTCACCACGTGCGGGCG---------------------- 4014787

Qy     539 nAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrPr 559
                          :::|||   |||         ||||||     :::
Db 4014786 ---------------GAAGGCAACCGGGCACTCGATGCCGGCGACCACGAC-------- 4014751

Qy     559 oGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLy 579
                   ||||||   |||   :::                |||           |||
Db 4014750 ----GGGGCCTTGCGAAGCGCTAGCGCGATTCGGGCGATGATG---------GGCATCCT 4014704

Qy     579 sGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLe 599
            |||   :::|||   |||     |||
Db 4014703 GGGCTGTGACCCGCTCGACCAGCGCTGG------------------------------ 4014676

Qy     599 uGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu-- 618
            |||   :::|||      |||        |||      :::    ::::::
Db 4014675 -GAATCCCGAGACGAAACCTCGGCAGCGCTGGCCGCCGTCGATGTGCTGGTCCAGGCTGA 4014617

Qy     619 -PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLe 638
               :::   :::          |||         |||       |||
Db 4014616 ACTACAGAATCGGGAAAAGGCCCGCGAGCAGCGCAACTGGGCGCTC------------- 4014571

Qy     638 uSerAsnAsnAspIleThrGly------GlyAsnSerGlySerProValPheAspLysAs 656
                  ::::::|||   |||              :::|||       |||   |||
Db 4014570 ----GCCGACGAGATCCGGGGTCGGCTGAAGAGAGCCGGCATCGAGGTCACCGAC----- 4014520

Qy     656 nGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGl 676
                             |||               |||      :::    ||||||   ::
Db 4014519 ---------------ACCGCCGAGGGGCCACAGTGGTCGCTGCTGGGTGGTGACACCAA 4014476

Qy     676 uPheGluProAspLeuGlnArg 683
             :      |||       |||
Db 4014475 GTAGATGCCCGGTAACTCTCGG 4014454
```

Search completed: May 23, 2003, 13:33:17
Job time : 4891.71 secs

```
                      GenCore version 5.1.4_p5_4578
            Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:         May 23, 2003, 10:12:58 ; Search time 186.87 Seconds
                                          (without alignments)
                                          5031.128 Million cell updates/sec

Title:          US-10-008-355-2
Perfect score:  3719
Sequence:       1 MQMKLKSILLGAALLLGASG..........LFMIDKWGQCPRLIQELKLI 712

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters:      1657494

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10008355/runat_16052003_110406_2649/app_query.fasta_1.1230
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10008355_@CGN_1_1_132_@runat_16052003_110406_2649
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published_Applications_NA:*
                1:  /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                10:  /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                12:  /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
                13:  /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                14:  /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
```

Result No.		Score	% Query Match	Length	DB	ID	Description
1		3719	100.0	2139	9	US-10-008-355-1	Sequence 1, Appli
2		133	3.6	2652	10	US-09-815-242-9284	Sequence 9284, Ap
3		127	3.4	2652	10	US-09-815-242-9586	Sequence 9586, Ap
4		116.5	3.1	15424	9	US-10-143-186-1	Sequence 1, Appli
5		116.5	3.1	15424	9	US-10-143-186-2	Sequence 2, Appli
6		116	3.1	3598	10	US-09-925-301-170	Sequence 170, App
7		113.5	3.1	2856	10	US-09-815-242-6368	Sequence 6368, Ap
8		112.5	3.0	10758	7	US-08-781-986A-221	Sequence 221, App
9		110.5	3.0	2082	9	US-09-769-787-197	Sequence 197, App
10		110.5	3.0	15450	9	US-10-127-391-1	Sequence 1, Appli
11		110	3.0	2319	9	US-10-121-032-14	Sequence 14, Appl
12		110	3.0	2319	9	US-10-093-037-14	Sequence 14, Appl
13		110	3.0	2538	9	US-10-027-806-61	Sequence 61, Appl
14		110	3.0	2538	9	US-10-034-623-61	Sequence 61, Appl
15		110	3.0	2538	9	US-10-027-801-61	Sequence 61, Appl
16		110	3.0	8930	9	US-10-192-584-1	Sequence 1, Appli
17	c	110	3.0	32998	9	US-10-027-806-1	Sequence 1, Appli
18	c	110	3.0	32998	9	US-10-034-623-1	Sequence 1, Appli
19	c	110	3.0	32998	9	US-10-027-801-1	Sequence 1, Appli
20		109	2.9	6702	9	US-09-769-787-209	Sequence 209, App
21		108	2.9	3825	10	US-09-974-300-1925	Sequence 1925, Ap
22		108	2.9	11871	10	US-09-070-927A-171	Sequence 171, App
23		107	2.9	38155	9	US-10-114-170-79	Sequence 79, Appl
24		106.5	2.9	1572	10	US-09-815-242-7820	Sequence 7820, Ap
25		106.5	2.9	5761	9	US-09-928-847B-26	Sequence 26, Appl
26		106	2.9	2472	9	US-10-100-049-7	Sequence 7, Appli
27		106	2.9	3358	9	US-10-100-049-20	Sequence 20, Appl
28		104.5	2.8	2610	10	US-09-815-242-6526	Sequence 6526, Ap
29		104.5	2.8	3513	9	US-09-938-842A-862	Sequence 862, App
30		104.5	2.8	5308	10	US-09-070-927A-444	Sequence 444, App
31		104.5	2.8	9821	7	US-08-781-986A-470	Sequence 470, App
32		104	2.8	3731	10	US-09-782-874-1	Sequence 1, Appli
33		104	2.8	7639	9	US-09-971-536-28	Sequence 28, Appl
34		104	2.8	14121	9	US-09-802-640-31	Sequence 31, Appl
35		104	2.8	14121	9	US-09-920-033-3	Sequence 3, Appli
36		103.5	2.8	1582	9	US-09-470-526-1	Sequence 1, Appli
37		103.5	2.8	1959	9	US-09-470-526-8	Sequence 8, Appli
38		103.5	2.8	3640	9	US-09-292-758-144	Sequence 144, App
39		103.5	2.8	4392	9	US-10-127-391-3	Sequence 3, Appli
40		103	2.8	1191	10	US-09-801-368-137	Sequence 137, App
41		103	2.8	2387	10	US-09-974-300-643	Sequence 643, App
42		102.5	2.8	4142	9	US-09-925-299-138	Sequence 138, App
43		102.5	2.8	4142	10	US-09-925-299-138	Sequence 138, App
44		102.5	2.8	6142	9	US-10-014-436-1	Sequence 1, Appli
45	c	102.5	2.8	7053	10	US-09-070-927A-78	Sequence 78, Appl

```
                          ALIGNMENTS

RESULT 1
US-10-008-355-1
; Sequence 1, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
;  APPLICANT: Travis, James
;  APPLICANT:  Potempa, Jan S
;  APPLICANT:  Banbula, Agnieszka
;  TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
;  FILE REFERENCE: 235.00440101
;  CURRENT APPLICATION NUMBER: US/10/008,355
;  CURRENT FILING DATE:  2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/246,827
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 2139
;   TYPE: DNA
;   ORGANISM: Porphyromonas gingivalis
US-10-008-355-1

Alignment Scores:
Pred. No.:               0              Length:         2139
Score:                   3719.00        Matches:        712
Percent Similarity:      100.00%        Conservative:   0
Best Local Similarity:   100.00%        Mismatches:     0
Query Match:             100.00%        Indels:         0
```

```
DB:                    9              Gaps:         0

US-10-008-355-2 (1-712) x US-10-008-355-1 (1-2139)

Qy        1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 ATGCAAATGAAATTAAAAAGTATTCTTCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60

Qy       21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAsp 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCCTCAACGAACTCAATCAGGAGAATCTGGAT 120

Qy       41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAGCCG 180

Qy       61 SerIleAlaAsnAlaValValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240

Qy       81 GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 CAGGGCCTGATCTTTACCAACCACCACTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300

Qy      101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuPro 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GTGGATCACGACTATCTGCGCGATGGTTTCGTTTCTCGCACGATGGGTGAGGAGCTTCCG 360

Qy      121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA 420

Qy      141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 GGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGC 480

Qy      161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 CAAGAACTGGCCAAAAAAGAAAATGCAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540

Qy      181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 TATTCCAACAACGAATACTTCCTCATCGTCTACGATGTATTCAAGGACGTTCGTATGGTA 600

Qy      201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 TTTGCTCCTCCCAGCTCTGTAGGTAAGTTCGGAGGCGATACGGACAACTGGATGTGGCCG 660

Qy      221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 CGTCACACGGGCGACTTCAGCGTATTCCGCGTGTATGCCGGTGCCGACAACCGGCCGGCC 720

Qy      241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln 260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 GAATACAGCAAGGACAATAAACCCTATAAGCCCGTTTACTTCGCTGCCGTATCCATGCAA 780

Qy      261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTAC 840

Qy      281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 CTCACTTCTTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAATCCTCGTATCGAAGTT 900

Qy      301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC 960

Qy      321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 AAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020

Qy      341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 CGCGGTCTCGCTCGTCTTGACGTGATAGGTCGTAAGCGTGCCGAGGAAAGAGCATTCGCA 1080

Qy      361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 GACTGGATCCGTAAGAACGGCAAGAGTGCTGTCTATGGCGATGTATTGTCTTCTCTCGAA 1140

Qy      381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 AAGGCTTATAAGGAAGGAGCCAAGGCCAACCGTGAGATGACTTATTTGAGCGAGACGCTC 1200

Qy      401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 TTCGGTGGTACCGAGGTGGTTCGTTTTGCACAGTTTGCCAACGCATTGGCTACAAATCCT 1260

Qy      421 AspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSer 440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 GATGCTCATGCCGGTATCCTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCCTCG 1320

Qy      441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAsp 460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 CTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGGCGTATCCCTGCCGAC 1380

Qy      461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 AAGCTCCCCGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440

Qy      481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 TATGCAGACTTCGTATTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCCATGCCATG 1500

Qy      501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 CTCAAGTCCATGGACAAGGAAAAGTTTGCCAAGGCTATCGAGAAAGATCCGGCAGTAGAG 1560

Qy      521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 CTTTCCAAGAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCCGATGCGATGGCCAATGCC 1620

Qy      541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 TATGCCATTGAGAAGGGCAAGCGTCTTTTCTTTGCCGGTTTGCGTGAGATGTACCCCGGA 1680

Qy      561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 CGTGCTCTGCCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740

Qy      581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 TATGAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGGCAAGGGCGTATTGGAG 1800

Qy      601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCGTACAGGAGAATATCCTCGACCTCTTCCGC 1860

Qy      621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 ACCAAAAACTATGGTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCCTATCGAAC 1920

Qy      641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 AACGACATCACGGGCGGTAACTCCGGTAGCCCCGTATTCGATAAGAACGGCCGTCTGATC 1980

Qy      661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 GGTCTTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040

Qy      681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGly 700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATTGACAAATGGGGT 2100

Qy      701 GlnCysProArgLeuIleGlnGluLeuLysLeuIle 712
            ||||||||||||||||||||||||||||||||||||
```

```
Db   2101 CAGTGCCCCCGTCTCATCCAAGAGCTGAAGTTGATC 2136

RESULT 2
US-09-815-242-9284
; Sequence 9284, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9284
;   LENGTH: 2652
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(2652)
US-09-815-242-9284

Alignment Scores:
Pred. No.:                0.00017      Length:          2652
Score:                    133.00       Matches:         106
Percent Similarity:       33.20%       Conservative:    64
Best Local Similarity:    20.70%       Mismatches:      164
Query Match:              3.58%        Indels:          178
DB:                       10           Gaps:            25

US-10-008-355-2 (1-712) x US-09-815-242-9284 (1-2652)

Qy    192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
          |||||||||                       |||      :::|||     |||
Db     76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 135

Qy    204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
          |||:::     ||||||                                |||
Db    136 CCAAACGTTACAGGTAAACTT-------------------------------CACCTT 162

Qy    224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
          |||                                                 ::::::
Db    163 GGT-------------------------------------------------CACGCT 171

Qy    244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
             |||             :::  :::                 |||||||||:::
Db    172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAAACGCATGCAAGGTTTTGAT 228

Qy    264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                              |||:::      ||||||               :::    |||

Db    229 -----------------ACCCTTTGGCTTCCTGGGATGGACCACGCAGGGATTGCCACT 270

Qy    283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                   ||||||:::|||:::      |||        |||     ::::::
Db    271 CAGGCTAAGGTTGAGGAGCGCTTGCGTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT 330

Qy    303 -----------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                     |||        ||||||:::                 :::   ||||||
Db    331 GAGTCTTTCTTGACGAAAGTCTGGGAATGGAAAGAC-----------GAATATGCCACT 378

Qy    319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
           ||||||              |||         |||    :::|||    :::
Db    379 ACCATCAAGGAACAATGGGGCAAGATGGGGCTCTCTGTAGACTATTCTCGTGAGCGTTTC 438

Qy    338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
             ::::::    |||||||:::                        :::|||    :::
Db    439 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 474

Qy    358 AlaPheAlaAspTrpIleArgLys-----------AsnGlyLysSerAlaValTyrGly 373
             |||   |||      :::|||              |||:::        :::
Db    475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 534

Qy    374 Asp---------------ValLeuSerSerLeuGluLysAlaTyrLys------------ 384
          |||                 ||||||   :::|||      :::|||
Db    535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTGGAAGGT 594

Qy    385 -----------------------------GluGlyAlaLysAlaAsnArgGluMetThr 394
                                       :::|||:::::::|||        |||
Db    595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGCCCTTGAAGTTGCTACA 654

Qy    395 TyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                   ||||||:::||||||
Db    655 ACTCGTCCTGAGACTATGTTTGGGGACGTT------------------------------ 684

Qy    415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
          |||:::|||   ||||||:::                               |||   :::|||
Db    685 GCGATTGCGGTTAATCCAGAA--------------------------GACCCGCGCTAC 717

Qy    435 LysAsp------------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
          ||||||                  ||||||   :::     |||:::::|||
Db    718 AAGGACTTGATTGGTAAAAACGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy    449 MetLeuAsp---------------------------------------IleValArgArg 455
          :::    |||                                     |||
Db    775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy    456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
             |||   |||   |||         ||||:::
Db    835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACATCATG 894

Qy    476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
             |||         :::   ||||||:::   |||   :::         |||
Db    895 AACGACGACGGAACTATGAACGAGCTTGTCTTTGAATTTTCAGGCATG---------GAT 945

Qy    496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
          :::|||   |||               |||         ||||||      |||
Db    946 CGTTTTGAAGCTCGT---------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy    516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla------------ 531
                ||||::::::   |||   |||   :::            :::
Db    991 GGTGCCCTTGTCAAAATCGAAAAACGTGTCCACAGTGTTGGTCACTCAGAACGTACAGGT 1050

Qy    532 ---------------------------------------------------IleGlnAla 534
                                                             :::
Db   1051 GTCGTAGTTGAGCCACGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy    535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
          :::|||:::||||||         :::|||
Db   1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG-------------------------- 1143

Qy    555 ArgGluMetTyrProGlyArgAlaLeuProSer--------------------------- 565
             |||   ||||||   |||         :::
Db   1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203
```

```
Qy     566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
               |||        :::   ::::::    :::|||                    |||
Db    1204 GACTGGGTTATCTCTCGTCAGCTCTGGTGGGGTCAC-----------------CAAATC 1245

Qy     586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
               |||||||||||            |||:::     :::   ::::::   |||:::
Db    1246 CCTGCTTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy     606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
               |||    :::        |||::::::||||||
Db    1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335
```

RESULT 3
US-09-815-242-9586
; Sequence 9586, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9586
; LENGTH: 2652
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2652)
US-09-815-242-9586

Alignment Scores:

Pred. No.:	0.000784	Length:	2652
Score:	127.00	Matches:	105
Percent Similarity:	33.01%	Conservative:	64
Best Local Similarity:	20.51%	Mismatches:	165
Query Match:	3.41%	Indels:	178
DB:	10	Gaps:	25

US-10-008-355-2 (1-712) x US-09-815-242-9586 (1-2652)

```
Qy     192 AspValPhe-----------------------LysAspValArgMetValPheAlaPro 203
               |||||||||                        |||       :::|||      |||
Db      76 GATGTTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTGATTCCACCA 135

Qy     204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
               |||:::      ||||||                                  |||
Db     136 CCAAACGTAACTGGGAAACTT-------------------------------CACCTT 162

Qy     224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
           |||                                                    ::::::
Db     163 GGT-----------------------------------------------------CACGCT 171

Qy     244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
              |||             :::   :::                   |||||||||:::
Db     172 TGGGATACG---ACTTTGCAGGATATCATCATCCGTCAAAAACGCATGCAGGGCTTTGAT 228

Qy     264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
                                |||:::     ||||||                 :::   |||
Db     229 -----------------ACCCTTTGGCTTCCAGGTATGGACCACGCTGGTATTGCGACT 270

Qy     283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
                       ||||||:::|||:::          |||         |||   ::::::
Db     271 CAGGCTAAGGTTGAGGAGCGCTTACGTGGTGAGGGTATTACGCGTTATGACCTAGGTCGT 330

Qy     303 -----------IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
                          |||          ||||||:::                :::   ||||||
Db     331 GAGTCTTTCTTGACGAAGGTCTGGGAATGGAAAGAC-----------GAATATGCCACT 378

Qy     319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
              ||||||          |||          |||   :::|||   :::
Db     379 ACCATCAAGGAACAATGGGGCAAGATGGGGCTTTCTGTAGACTATTCTCGTGAGCGTTTC 438

Qy     338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
              ::::::   ||||||:::                                 :::|||      :::
Db     439 ACTCTTGACGAAGGTTTGTCA-----------------------AAAGCTGTTCGTAAG 474

Qy     358 AlaPheAlaAspTrpIleArgLys-----------AsnGlyLysSerAlaValTyrGly 373
              |||   |||      :::|||                 |||:::        :::
Db     475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTATCGTGGTGAGTTTATCATCAACTGG 534

Qy     374 Asp--------------ValLeuSerSerLeuGluLysAlaTyrLys----------- 384
           |||              ||||||   :::|||      :::|||
Db     535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCACAAGGATGTAGAAGGT 594

Qy     385 -----------------------------GluGlyAlaLysAlaAsnArgGluMetThr 394
                                        :::|||::::::|||            |||
Db     595 GCCTTCTACCACATGAATTACATGCTGGAAGATGGTTCACGCGCCCTTGAAGTTGCTACA 654

Qy     395 TyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
                       ||||||:::||||||
Db     655 ACTCGTCCTGAGACTATGTTTGGGACGTT----------------------------- 684

Qy     415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
           |||:::|||   ||||||:::                                 |||   :::|||
Db     685 GCGGTTGCGGTTAATCCAGAA-------------------------GACCCGCGCTAC 717

Qy     435 LysAsp-----------------TyrLeuProSerLeuAspArgLysValLeuProAla 448
           ||||||                  ||||||    :::         |||::::::|||
Db     718 AAGGACTTGATTGGTAAAAACGTCATCCTTCCA---ATCGCTAATAAACTCATCCCAATC 774

Qy     449 MetLeuAsp---------------------------------------IleValArgArg 455
           :::   |||                                        |||
Db     775 GTTGGAGATGAGCACGCAGATCCTGAGTTTGGTACTGGTGTCGTGAAAATCACACCTGCC 834

Qy     456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
                 |||   |||   |||               |||:::
Db     835 CACGATCCAAATGACTTCTTGGTTGGCCAACGTCATAACTTGCCACAAGTCAACGTCATG 894

Qy     476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
              |||            :::      |||:::   |||      :::          |||
Db     895 AACGACGACGGAACTATGAATGAGCTTGCCTTTGAATTTTCAGGCATG---------GAT 945

Qy     496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
           :::|||   |||                  |||          ||||||      |||
Db     946 CGTTTTGAAGCTCGT---------------AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990

Qy     516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----------- 531
                       |||::::::   |||   |||   :::         :::
Db     991 GGTGCCCTTGTCAAAATCGAAAAACGTGTCCATTCAGTCGGTCACTCAGAACGGACAGGT 1050
```

```
Qy      532 ---------------------------------------------------IleGlnAla 534
                                                               :::
Db     1051 GTCGTAGTTGAGCCACGCTTGTCTACTCAATGGTTCGTCAAGATGGACCAATTGGCTAAG 1110

Qy      535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
            :::|||:::||||||          :::|||
Db     1111 AACGCCATTGCCAACCAAGACACAGAGGACAAG--------------------------- 1143

Qy      555 ArgGluMetTyrProGlyArgAlaLeuProSer--------------------------- 565
               |||   ||||||   |||         :::
Db     1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCCTTCAATGGATGGAAAATGTCCAC 1203

Qy      566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
            |||         :::   ::::::   :::|||                     |||
Db     1204 GACTGGGTTATCTCTCGTCAGCTTTGGTGGGGTCAC------------------CAAATC 1245

Qy      586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
               |||||||||||            |||:::    :::      ::::::   |||:::
Db     1246 CCTGCCTGGTACAAT------GCTGATGGTGAAATGTATGTCGGCGAAGAAGCTCCAGAA 1299

Qy      606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
               |||   :::         |||::::::||||||
Db     1300 GGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 1335
```

RESULT 4
US-10-143-186-1
; Sequence 1, Application US/10143186
; Publication No. US20030072771A1
; GENERAL INFORMATION:
; APPLICANT: MENGELING, WILLIAMS L.
; APPLICANT: VORWALD, ANN
; APPLICANT: LAGER, KELLY
; APPLICANT: ROOF, MIKE
; APPLICANT: BURKHART, KELLY
; APPLICANT: GORCYCA, DAVID E
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE,
; TITLE OF INVENTION: BASED ON ISOLATE JA-142
; FILE REFERENCE: 27093a
; CURRENT APPLICATION NUMBER: US/10/143,186
; CURRENT FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US/09/461,879
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 09/298,110
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 15424
; TYPE: DNA
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-10-143-186-1

Alignment Scores:

Pred. No.:	0.155	Length:	15424
Score:	116.50	Matches:	161
Percent Similarity:	33.12%	Conservative:	100
Best Local Similarity:	20.43%	Mismatches:	311
Query Match:	3.13%	Indels:	217
DB:	9	Gaps:	40

US-10-008-355-2 (1-712) x US-10-143-186-1 (1-15424)

```
Qy        7 SerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGlyValAlaLysAlaAspLys 26
            |||   :::|||    :::       ::::::|||         |||         |||   |||
Db    10035 TCTAATGTGTTGCGCAGCAGGTTCATCATCGGCCCGCCTGGT---------GCTGGGAAA 10085

Qy       27 GlyMetTrpLeuLeuAsnGluLeu------------------------------AsnGln 36
                  ||||||||    ::::::                              :::|||
Db    10086 ACATACTGGCTCCTTCAACAGGTCCAGGATGGTGATGCCATTTACACGCCAACTCACCAG 10145

Qy       37 GluAsnLeuAspArgMetArgGluLeuGly---------PheThrLeuProLeuAspSer 53
                  ||||||   :::|||   ||||||         |||   :::|||      :::
Db    10146 ACCATGCTCGATATGATTAGGGCTTTGGGGACGTGCCGGTTCAACGTCCCAGCAGGTACG 10205

Qy       54 LeuTyrSerPheAspLysProSerIleAlaAsn------AlaValValIlePheGlyGly 71
                     |||      ||||||                       :::::: ||||||
Db    10206 ACGCTGCAATTCCCTGCCCCCTCCCGTACCGGCCCTTGGGTTCGCATCCTAGCCGGCGGT 10265

Qy       72 GlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis--------- 88
               |||   |||                           :::   ||||||
Db    10266 TGGTGTCCTGGCAAGAATTCCTTCCTGGATGAAGCAGCGTATTGTAATCACCTTGATGTC 10325

Qy       88 ------------------------------------------------------------ 88

Db    10326 TTGAGGCTTCTTAGCAAAACTACCCTCACCTGTCTGGGAGATTTCAAACAACTCCACCCA 10385

Qy       89 ---------------HisCysGlyTyrGlyAlaIleGlnSerGlnSer------------ 99
                           ||||||            |||      |||:::
Db    10386 GTGGGTTTTGATTCTCATTGCTATGTTTTTGACATCATGCCTCAGACTCAACTGAAGACC 10445

Qy      100 ------------------------------ThrValAspHisAspTyrLeuArgAspGly 109
                                             :::   ||||||   ||||||
Db    10446 ATCTGGAGATTTGGACAGAATATCTGTGAGGCCATTCAGCCAGATTAC---AGGGACAAA 10502

Qy      110 PheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSerValLysTyrLeu 129
               ||||||                        :::            |||   |||:::
Db    10503 CTTGTATCCATG--------------------GTCAACACAACCCGTGTAACCTACGTG 10541

Qy      130 ArgLysIleValLysVal------------------------ThrAspLysValGluGly 141
               |||   ||||||                              |||:::   :::|||
Db    10542 GAAAAACCTGTCAAGTATGGGCAAGTCCTCACCCCTTACCACAGGGACCGAGAGGACGGC 10601

Qy      142 GlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGln 161
                        ||||||   :::      :::                  |||
Db    10602 ---------GCCATCACAATTGACTCCAGTCAAGGCGCCACATTTGATGTGGTTACACTG 10652

Qy      162 GluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyr 181
               |||      |||::::::   :::         :::      :::
Db    10653 CATTTGCCCACTAAAGATTCACTCAACAGGCAAAGAGCCCTTGTTGCT---------ATT 10703

Qy      182 SerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPhe 201
            :::         :::   :::   |||||||||      :::   ::::::   :::|||
Db    10704 ACCAGGGCAAGACATGCTGTCTTTGTGTATGACCCACACAGGCAACTGCAGAGCATGTTT 10763

Qy      202 AlaProProSer--------------------SerValGlyLysPheGlyGlyAspThr 214
                  |||:::                     :::|||   :::         |||
Db    10764 GATCTTCCTGCGAAAGGCACACCCGTCAACCTCGCTGTGCACCGT---------GACGAG 10814

Qy      215 AspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGly 234
                     :::      |||:::      :::   :::|||   :::
Db    10815 CAGCTGATCGTGCTAGATAGAAATAACAAAGAATGCACGGTTGCTCAGGCTCTAGGCAAT 10874

Qy      235 AlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPhe 254
               |||      |||               |||   :::         :::
Db    10875 GGGGATAAATTCAGGGCC---------ACAGACAAGCGCGTTGTAGATTCTCTCCGCGCC 10925

Qy      255 AlaAlaValSerMetGlnGlyTyrLysAla------AspAspTyrAlaMetThrIleGly 272
                        ::::::|||      :::               |||      :::|||
Db    10926 ATTTGTGCAGATCTGGAAGGGTCGAGCTCCCCGCTCCCCAAGGTCGCACACAACTTGGGA 10985

Qy      273 PheProGlySerThrAspArgTyrLeuThrSer--------------------------- 283
            |||      |||   |||      ||||||
Db    10986 TTTTATTTCTCGCCTGAT----TTGACACAGTTTGCTAAACTCCCGGTAGAACTTGCA 11039

Qy      284 ------TrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArg 301
                  |||   |||         :::   :::||||||   :::
Db    11040 CCCCACTGGCCCGTG---------GTGACAACCCAGAACAATGAAAAG------------ 11078

Qy      302 GlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGln-AlaThrArgIleLy 321
                              |||   :::   :::   |||       |||       |||
Db    11079 ------------------TGGCCAGACCGGTTGGTTGCTAGCCTTCGCCCCGTCCATAAG 11120

Qy      321 sTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLys-------------------- 334
            |||::::::      :::      |||   :::|||
Db    11121 TATAGCCGCGCGTGCATCGGTGCCGGCTACATGGTGGGCCCCTCAGTGTTTCTGGGCACC 11180
```

```
Qy      335 ------------AsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGl 350
                        ::::::|||      |||        |||||||||
Db    11181 CCTGGGGTTGTGTCATACTATCTCACAAAATTTGTCAGGGGCGAGGCTCAAATGCTTCCG 11240

Qy      350 yArgLysArg---------AlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGl 367
             |||:::            ||||||   |||       |||       |||      ||
Db    11241 GAGACAGTCTTCAGCACCGGCCGAATTGAGGTAGATTGCCGTGAGTATCTCGATGACCGG 11300

Qy      367 yLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAl 387
             |                     ::::::      ||||||     |||:::
Db    11301 GAGCGA------------GAAATTGCTGAGTCCCTCCCCCATGCTTTCATTGGCGACGT 11347

Qy      387 aLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVa 407
             |||                                 |||       ||||||        ||
Db    11348 CAAAGGC-------------------------ACTACCGTTGGAGGATGTCACCATGT 11380

Qy      407 l-----------------ArgPhe---------AlaGlnPheAlaAsnAlaLeuAlaTh 418
            |                 ||||||             |||                ::
Db    11381 CACCTCCAAATACCTTCCGCGCTTCCTTCCCAAGGAATCAGTCGCGGTAGTCGGGGTTTC 11440

Qy      418 rAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr----------- 434
             :::::|||         |||       :::   :::|||   |||   |||
Db    11441 AAGCCCCGGGAAAGCCGCAAAAGCAGTTTGCACATTAACAGATGTGTATCTCCCAGATCT 11500

Qy      435 -LysAspTyrLeu---ProSerLeuAspArgLysValLeuProAlaMetLeuAspIleVa 453
             :::   ||||||   |||               |||         |||||||||
Db    11501 CGAAGCTTACCTCCACCCAGAGACCCAGTCCAAGTGCTGGAAAATGATGTTGGACTTCAA 11560

Qy      453 lArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLy 473
                    |||:::      |||                 :::        :::      :::
Db    11561 GGAAGTTCGACTGATGGTCTGGAAGGACAAGACGGCCTATTTTCAACTTGAAGGCCGCCA 11620

Qy      473 sPheLysGly---AspThrLysLysTyrAlaAspPheValPheAspLysSerValValPr 492
             |||                   ||||||   ::::::                |||||
Db    11621 TTTCACCTGGTACCAGCTTGCAAGCTATGCCTCGTACATC-----------CGAGTTCC 11668

Qy      492 oTyrSerAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAl 512
             |   :::
Db    11669 TGTTAAC---------------------------------------------TCTAC 11680

Qy      512 aIleGluLysAspProAlaVal------GluLeuSerLysSerValIleAlaAlaAlaAr 530
             :::     ||||||   :::           |||               :::      :::
Db    11681 GGTGTATTTGGACCCCTGCATGGGCCCTGCCCTTTGCAACAGAAGAGTTGTCGGGTCCAC 11740

Qy      530 gAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPh 550
                       ||||||                |||       :::   |||   ::::::
Db    11741 TCATTGGGGAGCTGACCTCGCAGTCACCCCTTAT------GATTACGGTGCCAAAATCAT 11794

Qy      550 ePheAlaGlyLeu------ArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAs 568
               :::                ||||||   ||||||        :::    :::    |||
Db    11795 CCTGTCTAGTGCATACCATGGTGAAATGCCCCCTGGGTACAAAATCCTGGCGTGCGCGGA 11854

Qy      568 nPheThrMet------ArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGl 586
             |||::::::          :::   |||         |||:::|||          |||
Db    11855 GTTCTCGCTTGACGATCCAGTGAGGTACAAACACACCTGGGGGTTTGAA---TCGGATAC 11911

Qy      586 yAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSe 606
             |||    |||   |||        ||||||   |||       |||      :::      ::
Db    11912 AGCG---TATCTGTACGAGTTCACCGGAAACGGT------GAGGACTGGGAGGATTACAA 11962

Qy      606 rAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyAr 626
             :|||   |||      ::::::                    :::    |||   |||
Db    11963 TGATGCGTTTCGTGCGCGCCAG------------------AAAGGGAAAATTTATAAGGC 12004

Qy      626 gTyrAlaGluAsnGlyGlnLeuHisIle--------------------------AlaPh 637
               |||      :::    :::    |||
Db    12005 CACTGCCACCAGCATGAGGTTTCATTTTCCCCCGGGCCCTGTCATTGAACCAACTTTAGG 12064

Qy      637 eLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGl 657
             |||:::   |||::::::               |||          :::||||||||||   ||
Db    12065 CCTGAATTGAAATGAAATG---------GGGTCCATGCAAAGCCTCTTTGACAAAATTGG 12115

Qy      657 yArgLeuIleGlyLeuAlaPhe 664
             |:::|||              ||||||
Db    12116 CCAACTTTTCGTGGATGCTTTC 12137
```

RESULT 5
US-10-143-186-2
; Sequence 2, Application US/10143186
; Publication No. US20030072771A1
; GENERAL INFORMATION:
; APPLICANT: MENGELING, WILLIAMS L.
; APPLICANT: VORWALD, ANN
; APPLICANT: LAGER, KELLY
; APPLICANT: ROOF, MIKE
; APPLICANT: BURKHART, KELLY
; APPLICANT: GORCYCA, DAVID E
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE,
; TITLE OF INVENTION: BASED ON ISOLATE JA-142
; FILE REFERENCE: 27093a
; CURRENT APPLICATION NUMBER: US/10/143,186
; CURRENT FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US/09/461,879
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 09/298,110
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 15424
;   TYPE: DNA
;   ORGANISM: Porcine reproductive and respiratory syndrome virus
US-10-143-186-2

```
Alignment Scores:
Pred. No.:              0.155       Length:        15424
Score:                  116.50      Matches:       161
Percent Similarity:     33.12%      Conservative:  100
Best Local Similarity:  20.43%      Mismatches:    311
Query Match:            3.13%       Indels:        217
DB:                     9           Gaps:          40
```

US-10-008-355-2 (1-712) x US-10-143-186-2 (1-15424)

```
Qy        7 SerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGlyValAlaLysAlaAspLys 26
             |||   :::|||   :::      ::::::|||      |||         |||   |||
Db    10035 TCTAATGTGTTGCGCAGCAGGTTCATCATCGGCCCGCCTGGT---------GCTGGGAAA 10085

Qy       27 GlyMetTrpLeuLeuAsnGluLeu-----------------------------AsnGln 36
               |||||||||   ::::::                                 :::|||
Db    10086 ACATACTGGCTCCTTCAACAGGTCCAGGATGGTGATGTCATTTACACGCCAACTCACCAG 10145

Qy       37 GluAsnLeuAspArgMetArgGluLeuGly---------PheThrLeuProLeuAspSer 53
              ||||||   :::|||   ||||||         |||      :::|||        :::
Db    10146 ACCATGCTCGATATGATTAGGGCTTTGGGGACGTGCCGGTTCAACGTCCCAGCAGGTACG 10205

Qy       54 LeuTyrSerPheAspLysProSerIleAlaAsn------AlaValValIlePheGlyGly 71
                   |||      ||||||                        ::::::   ||||||
Db    10206 ACGCTGCAATTCCCTGCCCCCTCCCGTACCGGCCCTTGGGTTCGCATCCTAGCCGGCGGT 10265

Qy       72 GlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis--------- 88
               |||   |||                           :::   ||||||
Db    10266 TGGTGTCCTGGCAAGAATTCCTTCCTGGATGAAGCAGCGTATTGTAATCACCTTGATGTC 10325

Qy       88 ------------------------------------------------------------ 88

Db    10326 TTGAGGCTTCTTAGCAAAACTACCCTCACCTGTCTGGGAGATTTCAAACAACTCCACCCA 10385

Qy       89 --------------HisCysGlyTyrGlyAlaIleGlnSerGlnSer----------- 99
                          ||||||         |||         |||:::
Db    10386 GTGGGTTTTGATTCTCATTGCTATGTTTTTGACATCATGCCTCAGACTCAACTGAAGACC 10445

Qy      100 -----------------------------ThrValAspHisAspTyrLeuArgAspGly 109
                                          :::      ||||||   ||||||
Db    10446 ATCTGGAGATTTGGACAGAATATCTGTGATGCCATTCAGCCAGATTAC---AGGGACAAA 10502
```

```
Qy      110 PheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSerValLysTyrLeu 129
               ||||||                        :::            |||   |||:::
Db    10503 CTTGTATCCATG--------------------GTCAACACAACCCGTGTAACCTACGTG 10541

Qy      130 ArgLysIleValLysVal-----------------------ThrAspLysValGluGly 141
               |||  ||||||                               |||:::   :::|||
Db    10542 GAAAAACCTGTCAAGTATGGGCAAGTCCTCACCCCTTACCACAGGGACCGAGAGGACGGC 10601

Qy      142 GlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGln 161
                        |||||    :::      :::                  |||
Db    10602 ---------GCCATCACAATTGACTCCAGTCAAGGCGCCACATTTGATGTGGTTACACTG 10652

Qy      162 GluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyr 181
               |||      |||::::::   :::      :::      :::
Db    10653 CATTTGCCCACTAAAGATTCACTCAACAGGCAAAGAGCCCTTGTTGCT---------ATT 10703

Qy      182 SerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPhe 201
            :::         :::   :::   ||||||||       :::   ::::::   :::|||
Db    10704 ACCAGGGCAAGACATGCTATCTTTGTGTATGACCCACACAGGCAACTGCAGAGCATGTTT 10763

Qy      202 AlaProProSer--------------------SerValGlyLysPheGlyGlyAspThr 214
                  |||:::                        :::|||   :::         |||
Db    10764 GATCTTCCTGCGAAAGGCACACCCGTCAACCTCGCTGTGCACCGT---------GACGAG 10814

Qy      215 AspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGly 234
                     :::      |||:::         :::   :::|||      :::
Db    10815 CAGCTGATCGTGCTAGATAGAAATAACAAAGAATGCACGGTTGCTCAGGCTCTAGGCAAT 10874

Qy      235 AlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPhe 254
               |||         |||            |||   :::            :::
Db    10875 GGGGATAAATTCAGGGCC---------ACAGACAAGCGCGTTGTAGATTCTCTCCGCGCC 10925

Qy      255 AlaAlaValSerMetGlnGlyTyrLysAla------AspAspTyrAlaMetThrIleGly 272
                        ::::::|||      :::                  |||      :::|||
Db    10926 ATTTGTGCAGATCTGGAAGGGTCGAGCTCCCCGCTCCCCAAGGTCGCACACAACTTGGGA 10985

Qy      273 PheProGlySerThrAspArgTyrLeuThrSer-------------------------- 283
            |||   |||   |||         ||||||
Db    10986 TTTTATTTCTCGCCTGAT------TTGACACAGTTTGCTAAACTCCCGGTAGAACTTGCA 11039

Qy      284 ------TrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArg 301
                  |||   |||         :::      :::||||||      :::
Db    11040 CCCCACTGGCCCGTG---------GTGACAACCCAGAACAATGAAAAG------------ 11078

Qy      302 GlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGln-AlaThrArgIleLy 321
                              |||   :::   :::   |||       |||      |||
Db    11079 ------------------TGGCCAGACCGGTTGGTTGCTAGCCTTCGCCCCGTCCATAAG 11120

Qy      321 sTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLys-------------------- 334
                 |||::::::      :::      |||   :::|||
Db    11121 TATAGCCGCGCGTGCATCGGTGCCGGCTACATGGTGGGCCCCTCAGTGTTTCTGGGCACC 11180

Qy      335 ------------AsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGl 350
                         ::::::|||            |||      |||||||||
Db    11181 CCTGGGGTTGTGTCATACTATCTCACAAAATTTGTCAGGGGCGAGGCTCAAATGCTTCCG 11240

Qy      350 yArgLysArg---------AlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGl 367
             |||:::            ||||||   |||      |||      |||         ||
Db    11241 GAGACAGTCTTCAGCACCGGCCGAATTGAGGTAGATTGCCGTGAGTATCTTGATGACCGG 11300

Qy      367 yLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAl 387
             |                 ::::::         ||||||      |||:::
Db    11301 GAGCGA-------------GAAATTGCTGAGTCCCTCCCCCATGCTTTCATTGGCGACGT 11347

Qy      387 aLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVa 407
             |||                              |||      ||||||            ||
Db    11348 CAAAGGC--------------------------ACTACCGTTGGAGGATGTCACCATGT 11380

Qy      407 l-----------------ArgPhe---------AlaGlnPheAlaAsnAlaLeuAlaTh 418
             |                 ||||||                  |||          . ::
Db    11381 CACCTCCAAATACCTTCCGCGCTTCCTTCCCAAGGAATCAGTCGCGGTAGTCGGGGTTTC 11440

Qy      418 rAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr----------- 434
            :::|||          |||    :::   :::|||        |||   |||
Db    11441 AAGCCCCGGGAAAGCCGCAAAAGCAGTTTGCACATTAACAGATGTGTATCTCCCAGATCT 11500

Qy      435 -LysAspTyrLeu---ProSerLeuAspArgLysValLeuProAlaMetLeuAspIleVa 453
             :::   ||||||         |||                  |||            ||||||||
Db    11501 CGAAGCTTACCTCCACCCAGAGACCCAGTCCAAGTGCTGGAAAATGATGTTGGACTTCAA 11560

Qy      453 lArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLy 473
                   |||:::         |||            :::      :::      :::
Db    11561 GGAAGTTCGACTGATGGTCTGGAAGGACAAGACGGCCTATTTTCAACTTGAAGGCCGCCA 11620

Qy      473 sPheLysGly---AspThrLysLysTyrAlaAspPheValPheAspLysSerValValPr 492
             |||                     ||||||   ::::::                     |||||
Db    11621 TTTCACCTGGTACCAGCTTGCAAGCTATGCCTCGTACATC-----------CGAGTTCC 11668

Qy      492 oTyrSerAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAl 512
            |   :::
Db    11669 TGTTAAC---------------------------------------------TCTAC 11680

Qy      512 aIleGluLysAspProAlaVal------GluLeuSerLysSerValIleAlaAlaAlaAr 530
             :::      ||||||   :::         |||            :::      :::
Db    11681 GGTGTATTTGGACCCCTGCATGGGCCCTGCCCTTTGCAACAGAAGAGTTGTCGGGTCCAC 11740

Qy      530 gAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPh 550
                      ||||||                  |||   :::   |||   ::::::
Db    11741 TCATTGGGGAGCTGACCTCGCAGTCACCCCTTAT------GATTACGGTGCCAAAATCAT 11794

Qy      550 ePheAlaGlyLeu------ArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAs 568
                :::                  ||||||   ||||||      :::   :::   |||
Db    11795 CCTGTCTAGTGCATACCATGGTGAAATGCCCCCTGGGTACAAAATCCTGGCGTGCGCGGA 11854

Qy      568 nPheThrMet------ArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGl 586
             |||::::::         :::   |||               |||:::|||         |||
Db    11855 GTTCTCGCTTGACGATCCAGTGAGGTACAAACACACCTGGGGGTTTGAA---TCGGATAC 11911

Qy      586 yAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSe 606
             |||   |||   |||      ||||||   |||      |||      :::       . ::
Db    11912 AGCG---TATCTGTACGAGTTCACCGGAAACGGT------GAGGACTGGGAGGATTACAA 11962

Qy      606 rAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyAr 626
            :|||   |||   ::::::                        :::   |||   |||
Db    11963 TGATGCGTTTCGTGCGCGCCAG-----------------AAAGGGAAAATTTATAAGGC 12004

Qy      626 gTyrAlaGluAsnGlyGlnLeuHisIle--------------------------AlaPh 637
                |||   :::   :::   |||
Db    12005 CACTGCCACCAGCATGAGGTTTCATTTTCCCCCGGGCCCTGTCATTGAACCAACTTTAGG 12064

Qy      637 eLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGl 657
             |||:::   |||::::::               |||         :::|||||||||   ||
Db    12065 CCTGAATTGAAATGAAATG---------GGGTCCATGCAAAGCCTCTTTGACAAAATTGG 12115

Qy      657 yArgLeuIleGlyLeuAlaPhe 664
            |:::|||         ||||||
Db    12116 CCAACTTTTTGTGGATGCTTTC 12137
```

RESULT 6
US-09-925-301-170
; Sequence 170, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 170
;   LENGTH: 3598

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (16)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (22)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (964)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-170

Alignment Scores:
Pred. No.:              0.0203        Length:         3598
Score:                  116.00        Matches:        110
Percent Similarity:     33.80%        Conservative:   85
Best Local Similarity:  19.06%        Mismatches:     198
Query Match:            3.12%         Indels:         185
DB:                     10            Gaps:           30

US-10-008-355-2 (1-712) x US-09-925-301-170 (1-3598)

Qy      90 CysGlyTyrGlyAlaIleGlnSerGlnSerThrValAspHisAspTyrLeuArgAspGly 109
              ||||||:::|||                                       |||   |||
Db      39 TGTGGTTTCGGTGAG--------------------------------CTCTCAGACCGA 65

Qy     110 PheValSerArgThrMetGlyGluGluLeu---------------------------Pro 120
              |||:::::          |||      :::                        |||
Db      66 TTTCTAGCGTCCGTGCCGGGGACAGGTGTCAGAGGTCGRCTGCTGCAGACATGGCGGCCT 125

Qy     121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140
                 |||               :::   :::   |||:::   ||||||   :::|||
Db     126 CCACCGCG-GCCGGGAAGCAGCGGATTCCCAAAGTGGCCAAGGTGAAAAACAAAGCCCCG 184

Qy     141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160
              :::::::::   ||||||   |||      :::   |||||||:::|||:::|||
Db     185 GCTGAGGTACAG---ATAACTGCTGAA---CAACTCTTAAGAGAGGCTAAAGAAAGAGAA 238

Qy     161 GlnGluLeu-----------AlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIle 176
                 ||||||               :::::::::   ||||||   :::|||
Db     239 CTTGAGCTTCTTCCACCTCCACCTCAACAGAAGATCACAGATGAAGAAGAATTA------ 292

Qy     177 ValGluProPheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAsp 196
                                    |||:::|||   |||               |||:::|||
Db     293 --------------------AATGATTATAAACTAAGGAAAAGGAAGACTTTTGAAGAT 331

Qy     197 ValArgMetValPheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsn 216
                                         ::::::   |||                  |||
Db     332 --------------------------AATATAAGAAAAAACAGGACTGTGATTAGTAAC 364

Qy     217 TrpMet-----------TrpProArgHisThrGlyAspPheSerValPheArgValTyr 232
              |||:::            |||
Db     365 TGGATAAAATACGCACAATGG------------------------------------- 385

Qy     233 AlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProVal 252
                                    |||      |||:::     :::     :::   :::
Db     386 --------------------GAAGAAAGCCTAAAGGAGATTCAAAGGGCTCGATCCATA 424

Qy     253 TyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGly 272
              |||      |||:::   :::                  ||||||
Db     425 TACGAGCGTGCTTTAGATGTA------------------GACTAC--------------- 451

Qy     273 PheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsn 292
                                 |||   :::|||.  |||                 :::|||
Db     452 -----------------CGAAATATTACACTCTGGCTGAAATACGCAGAAATGGAAATG 493

Qy     293 GluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMet 312
              :::|||                  |||                  ||||||      |||:::
Db     494 AAGAATCGCCAAGTCAMCCATGCTCGA-----------AATATCTGGGACCGGGCCATA 541

Qy     313 SerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyr 332
                                 |||   :::         :::::::::                |||
Db     542 ---------------ACAACGCTGCCTCGAGTTAATCAGTTCTGGTACAAGTACACGTAC 586

Qy     333 TrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLys 352
               :::      :::|||                            :::|||   |||   :::
Db     587 ATGGAGGAAATGTTGGGA-----------------------AACGTTGCCGGTGCCCGG 622

Qy     353 ArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyr 372
              :::      ||||||        |||:::                  :::         :::
Db     623 CAGGTGTTTGAGCGC-----------TGGATGGAGTGGCAGCCTGAGGAGCAAGCCTGG 670

Qy     373 GlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGlu 392
                    :::   :::   |||      |||||||||                :::|||
Db     671 CACTCCTACATC---AACTTTGAGCTGAGATACAAAGAG---------GTGGATCGGGCC 718

Qy     393 MetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluVal---------ValArgPhe 409
               |||                  :::         :::|||          :::::::::
Db     719 CGCACCATTTATGAGCGAKTTGTCCTCGTGCACCCTGATGTTAAGAACTGGATCAAGTAT 778

Qy     410 AlaGlnPheAlaAsnAlaLeuAla------------------------------------ 417
              |||:::|||         |||
Db     779 GCCCGCTTTGAAGAAAAACATGCTTATTTGCCCATGCACGGAAAGTGTATGAGAGAGCT 838

Qy     418 -----------------ThrAsnProAspAlaHisAlaGlyIle---LeuLysSerLeu 430
                                :::   |||   |||         :::
Db     839 GTGGAATTCTTTGGAGATGAACATATGGATGAGCACCTTTATGTTGCCTTTGCCAAGTTT 898

Qy     431 AspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeu 450
              ::::::      |||::::::                     |||   :::
Db     899 GAAGAAAATCAGAAAGAGTTT------GAAAGGGTACGAGTGATTTACAAGTATGCCCTG 952

Qy     451 AspIleValArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsn------ 468
              |||               ||||||               ::::::|||||||||
Db     953 GAC------------AGAATTTCNAAACAAGATGCCCAAGAACTCTTTAAAAATTATACC 1000

Qy     469 ValIleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPheValPheAspLys 488
              :::   :::|||||||||   ||||||   :::            |||   :::      |||
Db    1001 ATCTTTGAGAAGAAGTTT---GGTGATAGGCGGGGTATTGAAGATATCATTGTGAGCAAA 1057

Qy     489 SerValValProTyrSerAspLysPheHisAlaMetLeuLysSerMetAspLys---Glu 507
                 |||   ::::::      |||               :::   |||
Db    1058 CGGAGATTCCAGTACGARGAAGAAGTGAAGGCGAATCCACACAATTATGATGCATGGTTT 1117

Qy     508 LysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLysSerValIleAla 527
              :::   :::   :::|||   |||      |||
Db    1118 GATTACTTGCGCTTGGTAGAAAGTGACGCAGAAGCTGAA-------------------- 1156

Qy     528 AlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLys 547
              |||   |||   :::         |||:::||||||         |||::::::   :::
Db    1157 GCCGTGAGAGAAGTCTATGAAAGGGCCATTGCCAATGTCCCACCCATTCAGGAGAAGAGG 1216

Qy     548 ArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAla 567
                                   |||   :::|||
Db    1217 CAC------------TGGAAGCGCTACATTTAT------------------------- 1237

Qy     568 AsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAla 587
Db    1238 ---------------------------------------------------------CTT 1240

Qy     588 TrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAsp 607
              |||   ||||||         :::      ||||||   :::|||||||:::
Db    1241 TGGATCAACTATGCACTCTATGAAGAA------TTGGAGGCAAAGGATCCTGAGAGGACA 1294

Qy     608 GluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyr 624
                    |||   :::   |||:::|||         |||   :::
Db    1295 AGACAGGTGTATCAAGCCTCT---TTGGAACTAATTCCTCACAAAAAGTTC 1342

RESULT 7
US-09-815-242-6368
; Sequence 6368, Application US/09815242
```

```
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6368
;   LENGTH: 2856
;   TYPE: DNA
;   ORGANISM: Escherichia coli
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(2856)
US-09-815-242-6368

Alignment Scores:
Pred. No.:               0.0272        Length:          2856
Score:                   113.50        Matches:         118
Percent Similarity:      30.56%        Conservative:    69
Best Local Similarity:   19.28%        Mismatches:      144
Query Match:             3.05%         Indels:          282
DB:                      10            Gaps:            30

US-10-008-355-2 (1-712) x US-09-815-242-6368 (1-2856)

Qy     150 MetGluArgLeuArgLysAlaGlnGluValCysGlnGluLeuAlaLys------------ 165
           ||||||:::            |||:::::::   |||   |||    :::
Db       1 ATGGAAAAGACATATAACCCACAAGATATCGAACAGCCGCTTTACGAGCACTGGGAAAAG 60

Qy     166 -----------LysGluAsnAlaAspGluAsnGln------LeuCysIleValGluPro 179
                      |||   |||   ||||||:::|||         ||||||:::   |||
Db      61 CAGGGCTACTTTAAGCCTAATGGCGATGAAAGCCAGGAAAGTTTCTGCATCATGATCCCG 120

Qy     180 PheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMet 199

Db     120 ------------------------------------------------------------ 120

Qy     200 ValPheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrp 219
                    ||||||
Db     121 ---------CCGCCG--------------------------------------------- 126

Qy     220 ProArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgPro 239
                                  :::|||   |||   :::|||               |||
Db     127 --------------------AACGTCAC-CGGCAGTTTGCATATGGGTCACGCCTTCCA 164

Qy     240 AlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMet 259
           |||   :::         :::   |||                                 :::
Db     165 GCAAACCATCATGGATACCATGATCCG-----------------------------CTA 194

Qy     260 GlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArg 279
                       :::         :::   :::         :::         ||||||
Db     195 TCAGCGCATGCAGGGCAAAAACACCCTGTGGCAGGTCGGTACTGACCACGCCGGGATCGC 254

Qy     280 TyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnPro---ArgIle 298
            |||               :::|||                  ||||||   |||:::
Db     255 TACCCAGATGGTCGTTGAGCGCAAGATTGCCGCAGAAGAAGGTAAAACCCGTCACGACTA 314

Qy     299 GluValArg--------------------GlyIleLysGlnGlyIleTrpLysGluAla 311
                 |||                    |||::::::   |||||||||:::
Db     315 CGGCCGCGAAGCTTTCATCGACAAAATCTGGGAATGGAAAGCGGAATCTGGCGGCACCAT 374

Qy     312 MetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGln-SerAlaAs 331
                 :::|||      :::|||                      |||  |||    ::
Db     375 TACCCGTCAGATGCGCCGTCTCGG----------------------CAACTCCGTCGA 410

Qy     331 nTyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyAr 351
            ::::   :::            |||:::   ||||||:::
Db     411 CTGGGAGCGTGAACGCTTCACCATGGACGAAGGCCTGTCCAAT-----------GCGGT 458

Qy     351 gLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaVa 371
             |||         |||   :::   :::   :::         ||||||      ||
Db     459 GAAAGAAGTTTTCGTTCGTCTGTATAAAGAAGACCTGATTTACCGTGGCAAACGCCTGGT 518

Qy     371 lTyrGlyAsp--------------ValLeuSerSerLeuGluLysAlaTyrLysGlu-- 385
           |      |||                  :::|||   ||||||            :::|||
Db     519 AAACTGGGATCCGAAACTGCGCACCGCTATCTCTGACCTGGAAGTCGAAAACCGCGAATC 578

Qy     386 ---------------------------------------GlyAlaLysAlaAsnArgGl 392
                                                  |||||||||
Db     579 GAAAGGTTCGATGTGGCACATCCGCTATCCGCTGGCTGACGGTGCGAAAACCGCAGACGG 638

Qy     392 uMetThrTyrLeu-----------------SerGluThrLeuPheGlyGlyThrGluVa 406
                 ||||||                   |||||||||   |||   |||   ||
Db     639 TAAAGATTATCTGGTGGTCGCGACTACCCGTCCAGAAACCCTGCTGGGCGATACTGGCGT 698

Qy     406 lValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIl 426
           |                                |||   ||||||:::
Db     699 A-----------------------------GCCGTTAACCCGGAA-------------- 714

Qy     426 eLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLe 446
                          |||   :::|||||||||   :::
Db     715 ------------GATCCGCGTTACAAAGATCTGATTGGCAAATAT-------------- 747

Qy     446 uProAlaMetLeuAspIleValArgArgArgIlePro------AlaAspLysLeuProAs 464
                 :::|||   :::|||   ||||||||||||       |||:::        ||
Db     748 ----GTCATTCTGCCGCTGGTTAACCGTCGTATTCCGATCGTTGGCGACGAACACGCCGA 803

Qy     464 pIlePheLysAsnVal------------------------------------------- 469
           |:::   |||
Db     804 CATGGAAAAAGGCACCGGCTGCGTGAAAATCACTCCGGCGCACGACTTTAACGACTATGA 863

Qy     470 -IleAspLysLys-----------------------------PheLysGlyAspThrLy 479
            :::   |||:::                              |||   ||||||   ::
Db     864 AGTGGGTAAACGTCACGCCCTGCCGATGATCAACATCCTGACCTTTGACGGCGATATCCG 923

Qy     479 sLysTyrAlaAspPheValPheAsp------------LysSerValValProTyrSerAs 495
            ::::   |||      |||||||||             :::|||   |||   ||||||
Db     924 TGAAAGCGCCCAG---GTGTTCGATACCAAAGGTAACGAATCTGACGTT---TATTCCAG 977

Qy     495 pLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLy 515
            :::     |||         :::   :::         |||:::||||||
Db     978 CGAAATCCCTGCAGAGTTCCAGAAACTG------GAGCGTTTTGCT------------- 1017

Qy     515 sAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----------- 531
                                  |||:::|||:::||||||      |||
Db    1018 ----------------GCACGTAAAGCAGTCGTTGCCGCAGTTGACGCGCTTGGCCTGCT 1061

Qy     531 ------------------------------------------------------------ 531

Db    1062 GGAAGAAATTAAACCGCACGACCTGACCGTTCCTTACGGCGACCGTGGCGGCGTAGTTAT 1121
```

```
Qy    532 ---------------------------IleGlnAlaAspAlaMetAlaAsn------Al 540
                                     ::::::||||||    :::|||
Db   1122 CGAACCAATGCTGACCGACCAGTGGTACGTGCGTGCCGATGTCCTGGCGAAACCGGCGGT 1181

Qy    540 aTyrAlaIleGluLysGly--------------------------LysArgLeuPhePh 551
              |||:::|||   |||                           :::   ::::::||
Db   1182 TGAAGCGGTTGAGAACGGCGACATTCAGTTCGTACCGAAGCAGTACGAAAACATGTACTT 1241

Qy    551 eAlaGlyLeu--------------------------ArgGluMetTyrProGlyArgAl 562
           |:::   :::                           |||:::::::::   |||
Db   1242 CTCCTGGATGCGCGATATTCAGGACTGGTGTATCTCTCGTCAGTTGTGGTGGGGTCACCG 1301

Qy    562 aLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGl 582
           :::|||
Db   1302 TATCCCG----------------------------------------------------- 1308

Qy    582 uProGlnAspGlyAlaTrpTyrAsn----------------------------------- 590
                      |||||||||:::
Db   1309 ------------GCATGGTATGACGAAGCGGGTAACGTTTATGTTGGCCGCAACGAAGA 1355

Qy    591 ------------TyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSe 606
                      :::       |||       |||:::    :::||||||
Db   1356 CGAAGTGCGTAAAGAAAATAACCTCGGTGCTGATGTTGTCCTGCGTCAGGAC-------- 1407

Qy    606 rAspGluPheAlaValGlnGluAsnIleLeuAsp 617
                            |||::::::||||||
Db   1408 ------------------GAAGACGTTCTCGAT 1422
```

RESULT 8
US-08-781-986A-221
; Sequence 221, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 221:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10758 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-221

```
Alignment Scores:
Pred. No.:               0.251        Length:        10758
Score:                   112.50       Matches:       100
Percent Similarity:      35.39%       Conservative:  78
Best Local Similarity:   19.88%       Mismatches:    168
Query Match:             3.03%        Indels:        158
DB:                      7            Gaps:          21
```

US-10-008-355-2 (1-712) x US-08-781-986A-221 (1-10758)

```
Qy    246 AsnLysProTyrLysProValTyrPheAlaAlaVal---SerMetGlnGly--------- 261
              ||||||   |||:::      |||:::    :::|||    ::::::||||||
Db    341 AACAAGGACTACGAAGGTGCTTATTATCCAAGTGTAGAAAATATCCAAGGTGCGAAAGAT 400

Qy    262 ----------------------------------TyrLysAlaAspAspTyrAlaMet 269
                                             |||   |||||||||    |||
Db    401 TATCATGAAGGAAAAACTGATCATATTAGCGGATTGAAGAAAATAGATGACTACACTATG 460

Qy    270 ThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrp-------------- 284
             :::    |||           :::    ||||||||     :::
Db    461 CAGGTTACATTTGATAAAAAACAAGAAAATTACTTAACAGGATTTATTACTGGACCTTTA 520

Qy    285 --------------------------GlyValGluAspArgIleGluAsnGluAsnAsn 295
                                    ::::::|||      :::::::::
Db    521 TTAAGTAAAAAATATTTATCAGATGTACCAATTAAAGATTTAGCGAAATCAGATAAAATC 580

Qy    296 ProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAsp 315
             :::      :::    ||||||             |||:::          :::
Db    581 CGAAAATATCCTATT---GGTATTGGACCGTATAAAGTTAAGAAAATCGTTCCAGGTGAG 637

Qy    316 GlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsn 335
                     :::|||:::                        :::||||||
Db    638 GCTGTTCAACTCGTTAAATTT--------------------GATGATTATTGG------ 670

Qy    336 SerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGlu 355
                         :::|||        ||||||    |||        |||    :::
Db    671 --------------CAAGGTAAGCCTGCACTAGACAAAATCAATTTAAAAGTTATTGAT 715

Qy    356 GluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspVal 375
          :::                                                      :::
Db    716 CAA-----------------------------------------------GCGCAAATT 727

Qy    376 LeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyr 395
          :::    ::::::||||||
Db    728 ATTAAGGCAATGGAAAAA------------------------------------------ 745

Qy    396 LeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAla 415
                            |||    :::|||          |||          |||
Db    746 -----------------GGCGATATTGATGTTGCGAATGATGCT------ACCGGTGCA 781

Qy    416 LeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLys 435
          :::|||   :::       :::             ||||||    |||
Db    782 ATGGCAAAAGATGCTAAGTCATCTAATGCTGGTCTCAAGGTATTATCT----------- 829

Qy    436 AspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArg 455
                   ||||||||||||       ::::::   :::    |||    |||    :::
Db    830 ------GCGCCAAGCTTAGACTACGGTTTAATAGGTTCGTATCTCATGATTACGATAAAA 883

Qy    456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
          ::::::          ||||||          :::|||:::    ||||||:::    :::
Db    884 AAGCTA------ATAAAACTGGTAAAGTGAGACCAAAATATGAA-GACAAAGAATTACGT 936

Qy    476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495

Db    936 ------------------------------------------------------------ 936

Qy    496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
                   |||||||||    ::::::|||:::||||||:::    ||||||
Db    937 ------AAAGCAATGCTTTATGCAATTGATAGAGAAAAATGGATCAAAGCGTTTTTCAAT 990

Qy    516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAla--- 534
                |||   |||::::::   |||    :::    :::          |||   |||
Db    991 GGTTACGCTAGTGAAATCAAT---AGTTTTGTACCATCTATGCATTGGATAGCAGCCAAT 1047

Qy    535 ------AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAla 552
                |||                |||    ||||||    |||::::::
```

```
Db   1048 CCTAAGGACCTAAATGATTACAAATATGATCCTGAAAAAGCTAAAAAAATCTTAGATAAG 1107

Qy    553 --------------------------GlyLeuArgGluMetTyrProGlyArgAlaLeu 563
                                     |||   ||||||         |||
Db   1108 TTAGGTTATAAAGATAGAGATGGTGACGGATTTAGAGAAGATCCTAAAGGTAATAAATTT 1167

Qy    564 ProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluPro 583
                :::   ||||||      :::   :::   ||||||         :::||||||
Db   1168 ------GAGATTAACTTT------AAACATAATTCAGGTTCTAATCCTACTTTTGAACCA 1215

Qy    584 GlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAsp 603
          :::      |||                        |||         ||||||
Db   1216 AGAACTGCTGCGATA--------------------AAAGATTTCTGGGAAAAAGTTGGC 1254

Qy    604 ProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsn 623
             |||:::::          |||:::
Db   1255 TTGAAAACAAATGTGAAGTTAGTAGAA--------------------------------- 1281

Qy    624 TyrGlyArgTyrAlaGluAsnGlyGlnLeuHis------------------------Ile 635
          :::|||:::|||   |||:::         |||                        :::
Db   1282 TTCGGTAAATATAATGAAGACTTAGCAAAACATGCATCTAATACGATTCCTGTTTATATG 1341

Qy    636 AlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLys 655
             ::::::::::::      :::|||                  |||:::
Db   1342 CCATATATCACATCTTATTTCATGACGCGTGCTATCGGCGACAGACCTTTAGTCGTCCCG 1401

Qy    656 AsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIle 675
          :::   :::         |||||||||   ||||||:::               |||
Db   1402 CAT---CAATCTCAGAACTTAGCATTTATTGGTAACTTTGCAGAAACAGAGCGAGACACT 1458

Qy    676 GluPheGluProAspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMet 695
             |||      :::               :::::::::      |||   |||
Db   1459 GTATTTACAACAGAATATTCGGTTCGTACTGCCATGGAAGCTGTTTATCAATTACTAAAT 1518

Qy    696 IleAspLys 698
          ||||||:::
Db   1519 ATAGATCGT 1527
```

RESULT 9
US-09-769-787-197
; Sequence 197, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 197
;   LENGTH: 2082
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-09-769-787-197

```
Alignment Scores:
Pred. No.:              0.0365        Length:          2082
Score:                  110.50        Matches:         132
Percent Similarity:     35.47%        Conservative:    100
Best Local Similarity:  20.18%        Mismatches:      246
Query Match:            2.97%         Indels:          176
DB:                     9             Gaps:            30

US-10-008-355-2 (1-712) x US-09-769-787-197 (1-2082)

Qy    127 LysTyrLeuArgLysIleVal---------------LysValThrAspLysValGluGly 141
          :::|||:::::||||||||                 |||   |||   |||
Db    238 GAATATGTAAAAAAAATAGTGGGTGAGAGCTATGCAAAATCAACTAAAAAGCGACATACA 297

Qy    142 GlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGln 161
                      :::   :::|||:::      ::::::         |||
Db    298 ATTACTGTAGCTCTAGTTAACGAGTTGAACAACATTAAG------AACGAGTATTTGAAT 351

Qy    162 GluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyr 181
          ::::::  :::         :::      :::   |||:::      ::::::|||
Db    352 AAAATAGTTGAATCAACCTCAGAAAGCCAACTACAGATACTGATGATGGAGAGTCGATCA 411

Qy    182 SerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPhe 201
                :::|||
Db    412 AAAGTAGATGAA------------------------------------------------ 423

Qy    202 AlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArg 221
                      :::|||   ||||||      |||:::   :::
Db    424 -----------GCTGTGTCTAAGTTTGAAAAGGACTCATCTTCT------------TCG 459

Qy    222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgPro------ 239
             :::   |||   |||                              :::|||
Db    460 TCAAGTTCAGACTCTTCC---------------------------ACTAAACCGGAAGCT 492

Qy    240 AlaGluTyrSerLysAspAsnLysProTyrLysPro------------------------ 251
          ::::::   :::|||   |||||||||   :::|||
Db    493 TCAGATACAGCGAAGCCAAACAAGCCGACAGAACCAGGAGAAAAGGTAGCAGAAGCTAAG 552

Qy    252 ------ValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMet 269
                |||      |||         :::         |||   :::|||
Db    553 AAGAAGGTTGAAGAAGCTGAGAAAAAAGCCAAGGATCAAAAAGAAGAAGATCGTCGT--- 609

Qy    270 ThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArg 289
                   :::|||         |||   :::   |||               |||
Db    610 ------AACTACCCAACCATTACTTACAAAACGCTTGAACTTGAAATTGCTGAGTCCGAT 663

Qy    290 IleGlu--------------------------------AsnGluAsnAsnProArgIle 298
          :::|||                                      |||   ||||||
Db    664 GTGGAAGTTAAAAAAGCGGAGCTTGAACTAGTAAAAGTGAAAGCTAACGAACCTCGAGAC 723

Qy    299 GluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
          |||   :::   |||||||||                 :::               :::||||||
Db    724 GAGCAAAAA---ATTAAGCAAGCAGAAGCGGAAGTTGAGAGTAAACAGCTGAGGCTACA 780

Qy    319 ArgIleLys--------------------TyrAlaSerLysTyrAlaGlnSerAlaAsn 331
          |||:::|||                        |||   :::   |||:::
Db    781 AGGTTAAAAAAAATCAAGACAGATCGTGAAGAAGCAGAAGAAGAAGCTAAACGAAGAGCA 840

Qy    332 TyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArg 351
                |||         |||      :::|||   |||:::         |||   |||
Db    841 GATGCTAAAGAGCAA---GGTAAACCAAAGGGGCGGGCAAAACGAGGAGTTCCTGGAGAG 897

Qy    352 LysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaVal 371
                      ::::::         |||         |||:::         |||:::|||
Db    898 CTAGCAACACCTGATAAAAAAGAAAATGAT------GCGAAGTCTTCAGATTCTAGCGTA 951

Qy    372 TyrGlyAspValLeuSerSerLeuGlu---LysAlaTyrLysGluGlyAlaLysAlaAsn 390
             :::      |||   |||         |||      |||:::   |||:::|||
Db    952 GGTGAAGAAACTCTTCCAAGCCCATCCCTGAAACCAGAAAAAAAGGTAGCAGAAGCTGAG 1011

Qy    391 ArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAla 410
          :::::::::            :::                  |||      |||
Db   1012 AAGAAGGTTGAAGAAGCTAAGAAAAAAGCCGAGGATCAAAAAGAAGAAGATCGCCGTAAC 1071

Qy    411 GlnPheAlaAsnAlaLeuAlaThr------------------------AsnProAspAla 422
                   |||         |||                           :::   :::
Db   1072 TACCCAACCAATACTTACAAAACGCTTGAACTTGAAATTGCTGAGTCCGATGTGGAAGTT 1131

Qy    423 HisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAsp 442
                      |||:::      :::   ::::::   |||:::         |||      :::
Db   1132 AAAAAAGCGGAGCTTGAACTAGTAAAAGAGGAAGCTAAGGAA------CCTCGAAACGAG 1185

Qy    443 ArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeu 462
             ||||||         |||      ::::::      ::::::      |||   :::|||
```

```
Db   1186 GAAAAAGTTAAGCAAGCAAAAGCGGAAGTTGAGAGTAAAAAAGCTGAGGCTACAAGGTTA 1245

Qy    463 ProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAla 482
                                      :::|||  |||  |||  ||||||
Db   1246 --------------------------GAAAAAATCAAGACAGATCGTAAAAAAGCAGAA 1278

Qy    483 AspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLys 502
          :::                                        |||     |||     :::
Db   1279 GAA-----------------------------GAAGCTAAACGAAAAGCAGCAGAAGAA 1308

Qy    503 SerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer 522
             |||||||||  |||:::        :::      ||||||   :::
Db   1309 GATAAAGTTAAAGAAAAACCAGCTGAACAACCACAACCAGCGCCGGCTCCAAAAGCAGAA 1368

Qy    523 LysSerValIleAla---------------AlaAlaArgAlaIleGlnAlaAspAlaMet 537
          |||      |||                  |||       :::|||:::
Db   1369 AAACCAGCTCCAGCTCCAAAACCAGAGAATCCAGCTGAACAACCAAAAGCAGAAAAACCA 1428

Qy    538 AlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhe---------PheAlaGlyLeu 554
          |||:::    |||  |||      |||              :::    |||
Db   1429 GCTGATCAACAAGCTGAAGAAGACTATGCTCGTAGATCAGAAGAAGAATATAATCGCTTG 1488

Qy    555 ArgGluMetTyrProGlyArg------------AlaLeuProSerAspAlaAsnPheThr 570
             :::    |||   :::             |||   ||||||
Db   1489 ACTCAACAGCAACCGCCAAAAACTGAAAAACCAGCACAACCATCTACTCCAAAA------ 1542

Qy    571 MetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsn 590
                                  |||:::::::   :::::::|||   ||||||
Db   1543 ------------------------ACAGGCTGGAAACAAGAAAACGGTATGTGGTACTTC 1578

Qy    591 TyrHisThrThrGlyLys---------GlyValLeuGluLysGlnAspProLysSerAsp 607
           |||:::|||    |||          |||     |||:::
Db   1579 TACAATACTGATGGTTCAATGGCGACAGGATGGCTCCAA--------------------- 1617

Qy    608 GluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyr 627
                                                    |||   |||   :::
Db   1618 ---------------------------------------------AACAATGGCTCATGG 1632

Qy    628 ---------AlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAsp------ 642
                   ||||||  :::       :::|||  ||||||
Db   1633 TACTACCTCAACAGCAATGGCGCTATGGCGACAGGATGGCTCCAAAACAATGGTTCATGG 1692

Qy    643 --------------------------IleThrGly------GlyAsnSerGlySerPro 651
                                    ||||||       |||:::||||||
Db   1693 TACTATCTAAACGCTAATGGTTCAATGGCAACAGGATGGCTCCAAAACAATGGTTCATGG 1752

Qy    652 ValPhe---AspLysAsnGlyArgLeuIle------GlyLeuAlaPheAspGlyAsnTrp 668
             :::   :::  ||||||   :::          |||   ::::::|||:::|||
Db   1753 TACTACCTAAACGCTAATGGTTCAATGGCGACAGGATGGCTCCAATACAATGGCTCATGG 1812

Qy    669 GluAlaMetSerGlyAsp-IleGluPheGluProAspLeuGlnArgThrIleSerValAs 688
             ::::::  ::: :::::::::      |||         |||:::::
Db   1813 TACTACCTAAACGCTAATGGTTCAATGGCGACAGGATGGCTCCAATACAATGGCTCATGG 1872

Qy    688 pIleArgTyrValLeuPheMetIleAspLysTrpGlyGln 701
                     |||  :::|||     |||   |||
Db   1873 TACTACCTAAACGCTAATGGTGATA------TGGCGACAG 1906
```

RESULT 10
US-10-127-391-1
; Sequence 1, Application US/10127391
; Patent No. US20020172690A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Products Inc.
; TITLE OF INVENTION: AN INFECTIOUS cDNA CLONE OF NORTH AMERICAN PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME (PRRS) VIRUS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: PC10278A
; CURRENT APPLICATION NUMBER: US/10/127,391
; CURRENT FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: US/09/470,661A
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 15450
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: cDNA
; OTHER INFORMATION: corresponding to No. US20020172690Alth American Porcine
; OTHER INFORMATION: Reproductive And Respiratory Syndrome (PRRS) Virus
; OTHER INFORMATION: Genome.
US-10-127-391-1

```
Alignment Scores:
Pred. No.:               0.713       Length:        15450
Score:                   110.50      Matches:       158
Percent Similarity:      32.57%      Conservative:  99
Best Local Similarity:   20.03%      Mismatches:    314
Query Match:             2.97%       Indels:        219
DB:                      9           Gaps:          39

US-10-008-355-2 (1-712) x US-10-127-391-1 (1-15450)

Qy       7 SerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGlyValAlaLysAlaAspLys 26
             |||  :::|||  :::    ::::::|||     |||         |||  |||
Db   10016 TCTAATGTGTTGCGCAGCAGGTTCATCATCGGTCCGCCCGGT---------GCTGGGAAA 10066

Qy      27 GlyMetTrpLeuLeuAsnGluLeu-----------------------------AsnGln 36
              |||||||||   ::::::                              :::|||
Db   10067 ACATACTGGCTCCTTCAGCAGGTCCAGGATGGTGATGTCATTTACACACCGACTCACCAG 10126

Qy      37 GluAsnLeuAspArgMetArgGluLeuGly---------PheThrLeuProLeuAspSer 53
              ||||||   :::|||  ||||||            |||    :::|||       :::
Db   10127 ACCATGCTCGACATGATTAGGGCTTTGGGGACGTGCCGGTTCAACGTCCCAGCAGGTACA 10186

Qy      54 LeuTyrSerPheAspLysProSerIleAlaAsn------AlaValValIlePheGlyGly 71
                 |||       ||||||                     ::::::  ||||||
Db   10187 ACGCTGCAATTCCCTGCCCCCTCCCGTACCGGCCCGTGGGTTCGCATCCTGGCCGGCGGT 10246

Qy      72 GlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis--------- 88
              |||   |||                          :::  ||||||
Db   10247 TGGTGTCCTGGTAAGAATTCCTTCCTGGATGAAGCAGCGTATTGTAATCACCTTGATGTC 10306

Qy      88 ------------------------------------------------------------ 88

Db   10307 TTGAGGCTCCTTAGCAAAACCACCCTTACCTGTCTGGGAGACTTCAAACAACTCCACCCA 10366

Qy      89 ---------------HisCysGlyTyrGlyAlaIleGlnSerGlnSer------------ 99
                          ||||||         |||  |||:::
Db   10367 GTGGGTTTTGATTCTCATTGCTATGTTTTTGACATCATGCCTCAGACCCAGTTGAAGACC 10426

Qy     100 -----------------------------ThrValAspHisAspTyrLeuArgAspGly 109
                                        :::      ||||||   ||||||
Db   10427 ATCTGGAGATTCGGACAGAACATCTGTGATGCCATCCAACCAGATTAC---AGGGACAAA 10483

Qy     110 PheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSerValLysTyrLeu 129
              ||||||                           :::         |||   |||:::
Db   10484 CTTGTGTCCATG--------------------GTCAACACAACCCGTGTAACCTACATG 10522

Qy     130 ArgLysIleValLysVal-----------------------ThrAspLysValGluGly 141
              |||  ||||||                           |||:::   :::|||
Db   10523 GAAAAACCTGTCAAGTATGGGCAAGTCCTCACCCCTTACCACAGGGACCGAGGACGGC 10582

Qy     142 GlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGln 161
                    ||||||   :::         :::                  |||
Db   10583 ---------GCCATCACAATTGACTCCAGTCAAGGCGCCACATTTGATGTGGTTACACTG 10633

Qy     162 GluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyr 181
              |||      |||::::::   :::      :::      :::
Db   10634 CATTTGCCCACTAAAGATTCACTCAACAGGCAAAGAGCCCTTGTTGCT---------ATC 10684

Qy     182 SerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPhe 201
           :::         :::   :::   |||||||||      :::   ::::::   :::|||
Db   10685 ACCAGGGCAAGACATGCTATCTTTGTGTATGACCCACACAGGCAATTGCAGAGCATGTTT 10744
```

```
Qy      202 AlaProProSer--------------------SerValGlyLysPheGlyGlyAspThr 214
                |||:::                      :::|||   :::         |||
Db    10745 GATCTTCCTGCGAAGGGCACACCCGTCAACCTCGCAGTGCACCGT---------GATGAG 10795

Qy      215 AspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGly 234
                         :::    |||:::      :::   :::|||    :::
Db    10796 CAGCTGATCGTACTGGATAGAAATAATAAAGAATGCACAGTTGCTCAGGCTATAGGCAAC 10855

Qy      235 AlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPhe 254
               |||         |||         |||   :::               :::
Db    10856 GGAGATAAATTCAGGGCC---------ACCGACAAGCGCGTTGTAGATTCTCTCCGCGCC 10906

Qy      255 AlaAlaValSerMetGlnGlyTyrLysAla------AspAspTyrAlaMetThrIleGly 272
                      ::::::|||         :::                  |||      :::|||
Db    10907 ATTTGTGCTGATCTGGAAGGGTCGAGCTCCCCGCTCCCCAAGGTCGCACACAACTTGGGA 10966

Qy      273 PheProGlySerThrAspArgTyrLeuThrSer--------------------------- 283
             |||     |||   |||     ||||||
Db    10967 TTTTATTTCTCACCTGAT------TTGACACAGTTTGCTAAACTCCCGGTAGACCTTGCA 11020

Qy      284 ------TrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArg 301
                  |||   |||         :::      :::||||||   :::
Db    11021 CCCCACTGGCCCGTG---------GTGACAACCCAGAACAATGAAAAG------------ 11059

Qy      302 GlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGln-AlaThrArgIleLy 321
                                |||   :::   :::   |||      |||      |||
Db    11060 ------------------TGGCCGGATCGGCTGGTTGCCAGCCTTCGCCCTGTCCATAAG 11101

Qy      321 sTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLys-------------------- 334
               |||   :::                  |||   :::|||
Db    11102 TATAGCCGTGCGTGCATTGGTGCCGGCTATATGGTGGGCCCCTCGGTGTTTCTAGGCACC 11161

Qy      335 ------------AsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGl 350
                            :::::         |||         |||||||||
Db    11162 CCTGGGGTCGTGTCATACTACCTCACAAAATTTGTCAAGGGCGAGGCTCAAGTGCTTC-- 11219

Qy      350 yArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAl 370
             |||:::::::::           |||         ||||||   :::
Db    11220 -CGGAGACAGTCTTCAGCACCGGCCGAATTGAGGTGGATTGC-CGGGAGTATCTTGATGA 11277

Qy      370 aValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAs 390
                         :::|||      ||||||      |||:::            |||
Db    11278 CAGGGAGCGAGAAGTTGCTGAGTCCCTCCCACATGCCTTCATTGGCGACGTCAAAGGC-- 11335

Qy      390 nArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVal-------- 407
                                          |||      ||||||      |||
Db    11336 ------------------------ACCACCGTTGGGGGATGTCATCATGTCACCTCCAA 11370

Qy      408 ---------ArgPhe--------AlaGlnPheAlaAsnAlaLeuAlaThrAsnProAs 421
                         ||||||            |||         :::::::|||
Db    11371 ATACCTTCCGCGCTTCCTTCCCAAGGAATCAGTCGCGGTAGTCGGGGTTTCGAGCCCCGG 11430

Qy      421 pAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr-----------LysAspTy 437
                    |||      :::    :::|||   |||   |||              :::  ||
Db    11431 GAAAGCCGCAAAAGCAGTGTGCACATTGACGGATGTGTACCTCCCAGACCTTGAGGCCTA 11490

Qy.     437 rLeu---ProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgAr 456
             ||||   |||               |||               |||||||||          ||
Db    11491 CCTCCACCCAGAGACTCAGTCTAAGTGCTGGAAAGTTATGTTGGACTTCAAGGAAGTTCG 11550

Qy      456 gIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGl 476
             |:::      |||                  :::      :::      :::   |||
Db    11551 ACTGATGGTCTGGAAAGACAAGACGGCCTATTTCCAACTTGAAGGCCGCTATTTCACCTG 11610

Qy      476 y---AspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAs 495
                          ||||||   ::::::                     ||||||   :::
Db    11611 GTATCAGCTTGCAAGCTACGCCTCGTACATC------------CGTGTTCCTGTCAAC-- 11656

Qy      495 pLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLy 515
                                                                      :::
Db    11657 ---------------------------------------------TCCACGGTGTATCT 11670

Qy      515 sAspProAlaVal------GluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGl 533
             ||||||   :::               |||               :::    :::
Db    11671 GGACCCCTGCATGGGCCCTGCCCTTTGCAACAGAAGAGTTGTCGGGTCCACCCATTGGGG 11730

Qy      533 nAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGl 553
             ||||||                  |||      :::   |||    ::::::      :::
Db    11731 AGCTGACCTCGCAGTCACCCCTTAT------GATTACGGTGCTAAAATCATCTTGTCTAG 11784

Qy      553 yLeu------ArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMe 571
                          ||||||   ||||||         :::   :::    |||   |||:::::
Db    11785 CGCTTACCATGGTGAAATGCCTCCTGGATACAAGATTCTGGCGTGCGCGGAGTTCTCGCT 11844

Qy      571 t------ArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTy 589
             :             :::   |||             |||:::|||      |||   |||   ||
Db    11845 CGACGACCCAGTCAAGTACAAACACACCTGGGGTTTTGAA---TCGGATACAGCG---TA 11898

Qy      589 rAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPh 609
             |   |||         ||||||   |||               :::|||
Db    11899 TCTGTATGAGTTCACCGGAAACGGT------------GAGGAC----------------- 11929

Qy      609 eAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGl 629
                      |||:::         |||   ||||||   :::       |||:::      ::
Db    11930 -------TGGGAGGATTACAATGATGCGTTTCGTGCGCGCCAGAAAGGGAAAATTTATAA 11982

Qy      629 uAsnGlyGlnLeuHisIleAlaPhe----------------------------------- 637
             :                  :::   |||
Db    11983 GGCCACTGCTACCAGCATGAAGTTTTATTTCCCCCGGGCCCCGTCATTGAACCAACTTT 12042

Qy      638 ----LeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAs 656
                    |||:::      |||:::::               |||         :::|||||||||
Db    12043 AGGCCTGAATTGAAATGAAATG---------GGGTCTATACAAAGCCTCTTCGACAAAAT 12093

Qy      656 nGlyArgLeuIleGlyLeuAlaPhe 664
             |||:::|||          ||||||
Db    12094 TGGCCAGCTTTTTGTGGATGCTTTC 12118
```

RESULT 11
US-10-121-032-14
; Sequence 14, Application US/10121032
; Patent No. US20020155550A1
; GENERAL INFORMATION:
;   APPLICANT: Bylina, Edward J.
;   TITLE OF INVENTION: GLYCOSIDASE ENZYMES
;   NUMBER OF SEQUENCES: 72
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Gray Cary Ware & Freidenrich LLP
;   STREET: 4365 Executive Drive, Suite 1600
;   CITY: San Diego
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 92121
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: Windows95
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/121,032
;   FILING DATE: 09-Apr-2002
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/09/134,078
;   FILING DATE: 13-AUG-1998
;   APPLICATION NUMBER: 08/949,026
;   FILING DATE: 10-OCT-1997
;   APPLICATION NUMBER: 60/056,916
;   FILING DATE: 06-DEC-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Haile, Lisa A.
;   REGISTRATION NUMBER: 38,347
;   REFERENCE/DOCKET NUMBER: 09010/024002
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 858/677-1456

```
;          TELEFAX: 858/677-1465
;   INFORMATION FOR SEQ ID NO: 14:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2319 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: Genomic DNA
;        FEATURE:
;             NAME/KEY:  Coding Sequence
;             LOCATION:  1...2316
;        SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-121-032-14

Alignment Scores:
Pred. No.:               0.0486        Length:          2319
Score:                   110.00        Matches:         155
Percent Similarity:      33.12%        Conservative:    105
Best Local Similarity:   19.75%        Mismatches:      268
Query Match:             2.96%         Indels:          258
DB:                      9             Gaps:            42

US-10-008-355-2 (1-712) x US-10-121-032-14 (1-2319)

Qy      74 ThrGly---IleThrValSerAspGlnGlyLeuIlePheThrAsnHisHisCysGlyTyr 92
           |||||    :::           |||:::|||       |||:::::::     ||||||||
Db     174 ACAGGCAAGGTCGAACAAGGTGATCGAGGC------TTTTCTGACCAATCCTGTGGATAC 227

Qy      93 GlyAlaIleGlnSer---GlnSer-ThrValAspHisAspTyrLeuArgAspGlyPheVa 111
                          ::::::   |||    ||||||||
Db     228 GAAAAAGAAAGAACTCTTCAAGGTTACTGTTGAC------------------------- 261

Qy     111 lSerArgThrMetGlyGluGluLeuProIleProGlyLeuSer----------------- 125
                          |||:::|||:::|||:::       :::
Db     262 ------------GGAAAAGAGATTCCCGTCTCAAGAGTGGAAAAGGCCGATCCCACGGA 308

Qy     126 ---------ValLysTyrLeuArgLysIleValLysValThrAspLysValGluGlyGl 142
                           |||:::|||    ::::::      :::   |||   |||
Db     309 CATAGACGTGACGAACTACGTGAGAATCGTCCTTTCTGAATCCCTGAAAGAAGAAGACCT 368

Qy     142 nLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGl 162
               |||   :::          :::|||       :::  |||:::   :::          ||
Db     369 CAGAAAAGACGTGGAACTGATCATAGAAGGTTACAAACCGGCAAGAGTCATCATGATGGA 428

Qy     162 uLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyrSe 182
           |                                        |||::::::   :::|||
Db     429 G-------------------------------------ATCCTGGACGACTACTATTA 449

Qy     182 rAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPheAl 202
             :::   |||          ||||||         |||   :::       |||:::::
Db     450 CGATGGAGAG---CTCGGAGCCGTATATTCTCCAGAGAAGACGATATTCAGAGTCTGGTC 506

Qy     202 aProProSerSer--------------ValGlyLysPheGlyGlyAspThrAsp----- 215
           :|||   |||                 :::   |||   |||   ||||||:::
Db     507 CCCCGTTTCTAAGTGGGTAAAGGTGCTTCTCTTCAAAAACGGAGAAGACACAGAACCGTA 566

Qy     216 --------------------------AsnTrpMetTrpProArgHisThr---GlyAs 225
                                      |||   :::|||               |||||
Db     567 CCAGGTTGTGAACATGGAATACAAGGGAAACGGGGTCTGGGAAGCGGTTGTTGAAGGCGA 626

Qy     225 pPhe---SerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLy 244
           |         ||||||  :::|||                    :::       |||  ||
Db     627 TCTCGACGGAGTGTTCTACCTCTAT--------------CAGCTGGAAAACTACGGAAA 671

Qy     244 sAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAl 264
           |                 |||   |||     ||||||               |||
Db     672 GATCAGAACAACCGTCGATCCT---TATTCGAAAGCGGTT-----------TACGCAAA 716

Qy     264 aAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTr 284
            :::    :::       :::            :::                   ||
Db     717 CAACCAAGAGAGCGCCGTTGTGAATCTTGCCAGGACAAACCCAGAA---------GGATG 767

Qy     284 pGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGlyIleLy 304
           |                    ||||||:::         |||:::||||||       |||    ::
Db     768 G-----------------GAAAACGACAGGGGACCGAAAATCGAA------GGATACGA 803

Qy     304 sGlnGlyIleTrpLysGluAlaMetSerAlaAsp-----------GlnAlaThrArgIl 320
           :      |||      |||       ||||||              :::   :::   ::
Db     804 AGACGCGATAATCTATGAAATACACATAGCGGACATCACAGGACTCGAAAACTCCGGGGT 863

Qy     320 eLysTyrAlaSerLysTyr---AlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMe 339
           :|||            |||          :::             |||         |||::
Db     864 AAAAAACAAAGGCCTCTATCTCGGGCTCACCGAAGAAAACACGAAAGGACCGGGCGGTGT 923

Qy     339 tAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPh 359
           :        ||||||:::   |||       :::|||                        ||
Db     924 GACAACAGGCCTTTCGCACCTTGTGGAACTCGGTGTTACACACGTTCATATACTTCCTTT 983

Qy     359 eAlaAspTrp--------------IleArgLysAsnGlyLysSerAlaValTyr----- 372
           |   |||:::                :::   |||:::   :::          :::
Db     984 CTTTGATTTCTACACAGGCGACGAACTCGATAAAGATTT-CGAGAAGTACTACAACTGGG 1042

Qy     373 ------------------------------GlyAspValLeuSerSerLeuGluLysAl 382
                                          |||    :::|||:::       :::|||
Db    1043 GTTACGATCCTTACCTGTTCATGGTTCCGGAGGGCAGATACTCAACCGATCCCAAAAACC 1102

Qy     382 aTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGl 402
              |||    ::::::  |||                   |||:::|||    :::
Db    1103 CACACACGAGAATCAGAGAAGTCAAAGAAATGGTCAAAGCCCTTCACAAACACGGTATAG 1162

Qy     402 y---------GlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAs 419
                     |||     |||          ::::::           |||
Db    1163 GTGTGATTATGGACATGGTGTTCCCTCACACCTACGGTATAGGCGAACTCTCTGCGTTCG 1222

Qy     419 nProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLys----------------- 433
              |||           :::|||::::::    |||:::
Db    1223 ATCAGACGGTGCCGTACTACTTCTACAGAATCGACAAGACAGGTGCCTATTTGAACGAAA 1282

Qy     434 ---------------------------------TyrLysAspTyrLeuProSerLeuAs 442
                                             :::   :::   :::   |||
Db    1283 GCGGATGTGGTAACGTCATCGCAAGCGAAAGACCCATGATGAGAAAATTCATAGTCGATA 1342

Qy     442 pArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAla-AspLysL 462
             |||   :::|||          :::       ||||||||     |||:::
Db    1343 CCGTCACCTACTGGGTAAA------GGAGTATCACATAGACGGATTCAGGTTCGATCAGA 1396

Qy     462 euProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyr- 481
           ::          :::||||||||||||       :::    ::::::
Db    1397 TG--------------GGTCTCATCGACAAAAAGACAATGCTCGAAGTCGAAAGAGCTC 1441

Qy     482 ------------------------------------------------------AlaA 483
                                                                 |||
Db    1442 TTCATAAAATCGATCCAACTATCATTCTCTACGGCGAACCGTGGGGTGGATGGGGAGCAC 1501

Qy     483 spPheValPheAspLysSer-----------------ValValProTyrSerAspLysP 497
               |||    ||||||                     |||      ::::::|||:::|
Db    1502 CGATCAGGTTTGGAAAGAGCGATGTCGCCGGCACACACGTGGCAGCTTTCAACGATGAGT 1561

Qy     497 heHisAlaMetLeuLys---------------------------------------S 503
           ||        ::::::
Db    1562 TCAGAGACGCAATAAGGGGTTCCGTGTTCAACCCGAGCGTCAAGGGATTCGTCATGGGAG 1621

Qy     503 erMetAspLysGlu---------------------------------------------L 508
                 ||||||                                                |
Db    1622 GATACGGAAAGGAAACCAAGATCAAAAGGGGTGTTGTTGGAAGCATAAACTACGACGGAA 1681

Qy     508 ysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer--------------- 522
           ||       |||:::      ||||||       ::::::
Db    1682 AACTCATCAAAAGTTTCGCCCTTGATCCAGAAGAAACTATAAACTACGCAGCGTGTCACG 1741

Qy     523 ------------------LysSerValIleAlaAla-----------AlaArgAlaI 532
                             |||:::    :::||||||                 :::
Db    1742 ACAACCACACACTGTGGGACAAGAACTACCTTGCCGCCAAAGCTGATAAGAAAAAGGAAT 1801

Qy     532 leGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheA 552
                 :::   :::     ||||||          :::                 |||
```

```
Db    1802 GGACCGAAGAAGAACTGAAAAACGCCCAGAAACTGGCTGGTGCGATACTTCTCACTTCTC 1861

Qy     552 laGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetA 572
              |||:::      ::::::   |||:::                  ||||||
Db    1862 AAGGTGTTCCTTTCCTCCACGGAGGGCAGGACTTCTGCAGGACGACGAATTTCAACGACA 1921

Qy     572 rgMetSerTyrGly---------SerIleLysGlyTyrGluProGlnAspGlyAlaTrp- 588
                 ||||||             ||||||   |||::::::     :::
Db    1922 AC---TCCTACAACGCCCCTATCTCGATAAACGGCTTCGATTACGAAAGAAAACTTCAGT 1978

Qy     589 ----------TyrAsnTyrHisThrThrGlyLysGlyVal-----LeuGluLysGlnA 603
                     :::|||||||||              |||||||:::          |||   |||:::
Db    1979 TCATAGACGTGTTCAATTACCAC---------AAGGGTCTCATAAAACTCAGAAAAGAAC 2029

Qy     603 spProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysA 623
              |||                                                ||||||   ||||
Db    2030 ACCCT-------------------------------------GCTTTCAGGCTGAAAA 2050

Qy     623 snTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn------- 640
             ||            ||||||       :::   |||:::   ||||||
Db    2051 AC-----------GCTGAAGAGATCAAAAAACACCTGGAATTTCTCCCGGGCGGGAGAA 2098

Qy     641 -------------------AsnAspIleThrGlyGlyAsnSerGlySerPro---ValP 653
                              |||       ||||||:::                  |||
Db    2099 GAATAGTTGCGTTCATGCTTAAAGACCACGCAGGTGGTGATCCCTGGAAAGACATCGTGG 2158

Qy     653 heAspLysAsnGlyArgLeuIleGlyLeuAlaPhe---------AspGlyAsnTrp---- 668
                   ||||||   |||            :::          :::|||   |||
Db    2159 TGATTTACAATGGAAACTTAGAGAAGACAACATACAAACTGCCAGAAGGAAAATGGAATG 2218

Qy     669 -------------------------------------GluAlaMetSerGlyAspIleG 676
                                                 |||   :::   |||   ||||
Db    2219 TGGTTGTGAACAGCCAGAAAGCCGGAACAGAAGTGATAGAAACCGTCGAAGGAACAATAG 2278

Qy     676 luPheGluPro 679
             ||   :::|||
Db    2279 AACTCGATCCG 2289

RESULT 12
US-10-093-037-14
; Sequence 14, Application US/10093037
; Publication No. US20030078397A1
; GENERAL INFORMATION:
;  APPLICANT: Jay M. Short
;  APPLICANT:  Bylina, Edward
;  APPLICANT:  Swanson, Ronald V.
;  APPLICANT:  Mathur, Eric J.
;  APPLICANT:  Lam, David E.
;  TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF
;  FILE REFERENCE: 09010-024006
;  CURRENT APPLICATION NUMBER: US/10/093,037
;  CURRENT FILING DATE:  2002-03-06
;  PRIOR APPLICATION NUMBER: US 09/910,579
;  PRIOR FILING DATE: 2001-07-20
;  PRIOR APPLICATION NUMBER: US 09/134,078
;  PRIOR FILING DATE: 1998-08-13
;  PRIOR APPLICATION NUMBER: US 08/949,026
;  PRIOR FILING DATE: 1997-10-10
;  PRIOR APPLICATION NUMBER: US 60/056,916
;  PRIOR FILING DATE: 1996-12-06
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 2319
;   TYPE: DNA
;   ORGANISM: Thermotoga maritima
US-10-093-037-14

Alignment Scores:
Pred. No.:              0.0486       Length:          2319
Score:                  110.00       Matches:         155
Percent Similarity:     33.12%       Conservative:    105
Best Local Similarity:  19.75%       Mismatches:      268
Query Match:            2.96%        Indels:          258
DB:                     9            Gaps:            42

US-10-008-355-2 (1-712) x US-10-093-037-14 (1-2319)

Qy      74 ThrGly---IleThrValSerAspGlnGlyLeuIlePheThrAsnHisHisCysGlyTyr 92
           ||||||    :::               |||:::|||      |||:::::::     |||||||||
Db     174 ACAGGCAAGGTCGAACAAGGTGATCGAGGC------TTTTCTGACCAATCCTGTGGATAC 227

Qy      93 GlyAlaIleGlnSer---GlnSer-ThrValAspHisAspTyrLeuArgAspGlyPheVa 111
                 ::::::   |||     |||||||||
Db     228 GAAAAAGAAAGAACTCTTCAAGGTTACTGTTGAC-------------------------- 261

Qy     111 lSerArgThrMetGlyGluGluLeuProIleProGlyLeuSer----------------- 125
                          |||:::|||:::|||:::          :::
Db     262 ------------GGAAAAGAGATTCCCGTCTCAAGAGTGGAAAAGGCCGATCCCACGGA 308

Qy     126 ---------ValLysTyrLeuArgLysIleValLysValThrAspLysValGluGlyGl 142
                           |||:::|||   ::::::      :::   |||     |||
Db     309 CATAGACGTGACGAACTACGTGAGAATCGTCCTTTCTGAATCCCTGAAAGAAGAAGACCT 368

Qy     142 nLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGl 162
               |||   :::         :::|||      :::   |||:::   :::          ||
Db     369 CAGAAAAGACGTGGAACTGATCATAGAAGGTTACAAACCGGCAAGAGTCATCATGATGGA 428

Qy     162 uLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyrSe 182
           |                                                |||::::::      :::|||
Db     429 G------------------------------------ATCCTGGACGACTACTATTA 449

Qy     182 rAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPheAl 202
            :::      |||                  ||||||         |||      :::      |||:::::
Db     450 CGATGGAGAG---CTCGGAGCCGTATATTCTCCAGAGAAGACGATATTCAGAGTCTGGTC 506

Qy     202 aProProSerSer--------------ValGlyLysPheGlyGlyAspThrAsp----- 215
            :|||   |||                  :::   |||   |||   ||||||:::
Db     507 CCCCGTTTCTAAGTGGGTAAAGGTGCTTCTCTTCAAAAACGGAGAAGACACAGAACCGTA 566

Qy     216 --------------------------AsnTrpMetTrpProArgHisThr---GlyAs 225
                                      |||   :::|||                  |||||
Db     567 CCAGGTTGTGAACATGGAATACAAGGGAAACGGGGTCTGGGAAGCGGTTGTTGAAGGCGA 626

Qy     225 pPhe---SerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLy 244
           |         ||||||   :::|||                        :::            |||   ||
Db     627 TCTCGACGGAGTGTTCTACCTCTAT--------------CAGCTGGAAAACTACGGAAA 671

Qy     244 sAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAl 264
           |                   |||      |||         ||||||                  |||
Db     672 GATCAGAACAACCGTCGATCCT---TATTCGAAAGCGGTT------------TACGCAAA 716

Qy     264 aAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTr 284
            :::      :::      :::      :::                                   ||
Db     717 CAACCAAGAGAGCGCCGTTGTGAATCTTGCCAGGACAAACCCAGAA---------GGATG 767

Qy     284 pGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGlyIleLy 304
           |                     ||||||:::         |||:::||||||      |||   ::
Db     768 G-----------------GAAAACGACAGGGGACCGAAAATCGAA------GGATACGA 803

Qy     304 sGlnGlyIleTrpLysGluAlaMetSerAlaAsp-----------GlnAlaThrArgIl 320
            :        |||      |||         ||||||                  :::   :::   :::
Db     804 AGACGCGATAATCTATGAAATACACATAGCGGACATCACAGGACTCGAAAACTCCGGGGT 863

Qy     320 eLysTyrAlaSerLysTyr---AlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMe 339
            :|||                 |||         :::                  |||         |||::
Db     864 AAAAAACAAAGGCCTCTATCTCGGGCTCACCGAAGAAAACACGAAAGGACCGGGCGGTGT 923

Qy     339 tAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPh 359
            :      ||||||:::      |||            :::|||                        ||
Db     924 GACAACAGGCCTTTCGCACCTTGTGGAACTCGGTGTTACACACGTTCATATACTTCCTTT 983

Qy     359 eAlaAspTrp---------------IleArgLysAsnGlyLysSerAlaValTyr----- 372
           |      |||:::                  :::   |||:::   :::         :::
Db     984 CTTTGATTTCTACACAGGCGACGAACTCGATAAAGATTT-CGAGAAGTACTACAACTGGG 1042
```

```
Qy     373 ------------------------------GlyAspValLeuSerSerLeuGluLysAl 382
                                         |||   :::|||:::      :::|||
Db    1043 GTTACGATCCTTACCTGTTCATGGTTCCGGAGGGCAGATACTCAACCGATCCCAAAAACC 1102

Qy     382 aTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGl 402
                   |||     ::::::  |||                |||:::|||   :::
Db    1103 CACACACGAGAATCAGAGAAGTCAAAGAAATGGTCAAAGCCCTTCACAAACACGGTATAG 1162

Qy     402 y---------GlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAs 419
                     |||      |||            ::::::        |||
Db    1163 GTGTGATTATGGACATGGTGTTCCCTCACACCTACGGTATAGGCGAACTCTCTGCGTTCG 1222

Qy     419 nProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLys---------------- 433
              |||            :::|||::::::     |||:::
Db    1223 ATCAGACGGTGCCGTACTACTTCTACAGAATCGACAAGACAGGTGCCTATTTGAACGAAA 1282

Qy     434 ---------------------------------TyrLysAspTyrLeuProSerLeuAs 442
                                            :::   :::   :::   |||
Db    1283 GCGGATGTGGTAACGTCATCGCAAGCGAAAGACCCATGATGAGAAAATTCATAGTCGATA 1342

Qy     442 pArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAla-AspLysL 462
            |||   :::|||             :::      |||||||||       |||:::::
Db    1343 CCGTCACCTACTGGGTAAA------GGAGTATCACATAGACGGATTCAGGTTCGATCAGA 1396

Qy     462 euProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyr- 481
           ::                     :::||||||||||||           :::   ::::::
Db    1397 TG--------------GGTCTCATCGACAAAAAGACAATGCTCGAAGTCGAAAGAGCTC 1441

Qy     482 --------------------------------------------------------AlaA 483
                                                                   |||
Db    1442 TTCATAAAATCGATCCAACTATCATTCTCTACGGCGAACCGTGGGGTGGATGGGGAGCAC 1501

Qy     483 spPheValPheAspLysSer------------------ValValProTyrSerAspLysP 497
              |||   ||||||                      |||       ::::::|||:::|
Db    1502 CGATCAGGTTTGGAAAGAGCGATGTCGCCGGCACACACGTGGCAGCTTTCAACGATGAGT 1561

Qy     497 heHisAlaMetLeuLys---------------------------------------S 503
           ||       ::::::
Db    1562 TCAGAGACGCAATAAGGGGTTCCGTGTTCAACCCGAGCGTCAAGGGATTCGTCATGGGAG 1621

Qy     503 erMetAspLysGlu------------------------------------------L 508
                   ||||||                                          |
Db    1622 GATACGGAAAGGAAACCAAGATCAAAAGGGGTGTTGTTGGAAGCATAAACTACGACGGAA 1681

Qy     508 ysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer---------------- 522
            ||      |||:::         ||||||          ::::::
Db    1682 AACTCATCAAAAGTTTCGCCCTTGATCCAGAAGAAACTATAAACTACGCAGCGTGTCACG 1741

Qy     523 -------------------LysSerValIleAlaAla------------AlaArgAlaI 532
                              |||:::   :::||||||               :::
Db    1742 ACAACCACACACTGTGGGACAAGAACTACCTTGCCGCCAAAGCTGATAAGAAAAAGGAAT 1801

Qy     532 leGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheA 552
                 :::   :::   ||||||         :::                  |||
Db    1802 GGACCGAAGAAGAACTGAAAAACGCCCAGAAACTGGCTGGTGCGATACTTCTCACTTCTC 1861

Qy     552 laGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetA 572
             |||:::       ::::::   |||:::                  ||||||
Db    1862 AAGGTGTTCCTTTCCTCCACGGAGGGCAGGACTTCTGCAGGACGACGAATTTCAACGACA 1921

Qy     572 rgMetSerTyrGly---------SerIleLysGlyTyrGluProGlnAspGlyAlaTrp- 588
                 ||||||             ||||||   |||::::::   :::
Db    1922 AC---TCCTACAACGCCCCTATCTCGATAAACGGCTTCGATTACGAAAGAAAACTTCAGT 1978

Qy     589 ----------TyrAsnTyrHisThrThrGlyLysGlyVal------LeuGluLysGlnA 603
                     :::|||||||||         ||||||:::      |||   |||:::
Db    1979 TCATAGACGTGTTCAATTACCAC---------AAGGGTCTCATAAAACTCAGAAAAGAAC 2029

Qy     603 spProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysA 623
              |||                                      ||||||   ||||
Db    2030 ACCCT---------------------------------------GCTTTCAGGCTGAAAA 2050

Qy     623 snTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn------- 640
              ||               ||||||        :::   |||:::   ||||||
Db    2051 AC------------GCTGAAGAGATCAAAAAACACCTGGAATTTCTCCCGGGCGGGAGAA 2098

Qy     641 --------------------AsnAspIleThrGlyGlyAsnSerGlySerPro---ValP 653
                               |||      ||||||:::                  |||
Db    2099 GAATAGTTGCGTTCATGCTTAAAGACCACGCAGGTGGTGATCCCTGGAAAGACATCGTGG 2158

Qy     653 heAspLysAsnGlyArgLeuIleGlyLeuAlaPhe---------AspGlyAsnTrp---- 668
                  ||||||   |||                 :::        :::|||   |||
Db    2159 TGATTTACAATGGAAACTTAGAGAAGACAACATACAAACTGCCAGAAGGAAAATGGAATG 2218

Qy     669 ------------------------------------GluAlaMetSerGlyAspIleG 676
                                                |||   :::   |||   ||||
Db    2219 TGGTTGTGAACAGCCAGAAAGCCGGAACAGAAGTGATAGAAACCGTCGAAGGAACAATAG 2278

Qy     676 luPheGluPro 679
             ||   :::|||
Db    2279 AACTCGATCCG 2289
```

RESULT 13
US-10-027-806-61
; Sequence 61, Application US/10027806
; Patent No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCORP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 61
;   LENGTH: 2538
;   TYPE: DNA
;   ORGANISM: Cenarchaeum symbiosum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(2538)
US-10-027-806-61

Alignment Scores:

Pred. No.:	0.0555	Length:	2538
Score:	110.00	Matches:	121
Percent Similarity:	37.06%	Conservative:	88
Best Local Similarity:	21.45%	Mismatches:	233
Query Match:	2.96%	Indels:	123
DB:	9	Gaps:	26

US-10-008-355-2 (1-712) x US-10-027-806-61 (1-2538)

```
Qy      20 GlyValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlu------------LeuAsn 35
           |||||||||   :::|||      :::::::::::   :::                :::
Db     901 GGCGTGGCGGATTCCGACATACCCCTGTACATGATGCGGGATTCGGCCACGCTCCGGCAC 960

Qy      36 GlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyr 55
              :::||||||         |||         |||:::                ||||||
Db     961 GGCGTCCATCTGGACCTGTACAGGACC------TTCTCGAACAGGTCGTTCCAGCTGTAT 1014

Qy      56 SerPhe--------------------AspLysProSerIleAlaAsnAlaValValIle 68
           :::|||                     |||::::::   |||::::::
Db    1015 GCATTTGCGGCAAAGTATACAGATTACTCCCTGAACAGCGTGTCCAAGGCGATGCTCGGC 1074

Qy      69 PheGlyGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
                 |||         |||:::::::::   |||      |||      |||   :::
Db    1075 GAGGGCAAGGTCGATTATGGCGTGTCTCTCGGGGATCTCACTCTATACCAGACTGCAAAC 1134

Qy      89 HisCysGlyTyrGlyAla------IleGlnSerGlnSerThrValAspHisAspTyrLeu 106
           :::|||   :::   |||                  ||||||      ::::::   |||
Db    1135 TATTGCTATCATGACGCGCGCCTGACGCTGGAGCTTAGCACCTTTGGGAACGAGATACTG 1194
```

```
Qy     107 ArgAspGlyPheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSer--- 125
             |||    |||       :::          :::||||||     :::|||
Db    1195 ATGGACCTCCTGGTGGTGACCAGCAGGATTGCCCGGATGCCCATCGATGATATGTCCCGC 1254

Qy     126 ---------ValLysTyrLeuArgLysIleValLysValThrAspLysValGluGlyGln 142
                        :::::::::|||   ::::::          :::
Db    1255 ATGGGCGTCTCGCAGTGGATAAGGAGCCTGCTGTACTATGAGCACAGGCAGCGCAACGCG 1314

Qy     143 LeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGlu 162
           |||            ||||||:::|||      :::::::::|||:::|||        :::
Db    1315 CTGATACCCCGCAGGGACGAGCTGGAA------AAGAGGTCTCAACAGGTAAGCAACGAC 1368

Qy     163 LeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluPro--------- 179
                 |||:::                     |||   :::|||||||||
Db    1369 GCCGTAATCAAGGACAAAAAGTTCCGCGGTGGTCTC---GTAGTCGAGCCTGAAGAGGGC 1425

Qy     180 --------------------------PheTyrSerAsnAsnGluTyrPheLeuIleVal 190
                                     |||   |||            :::   |||
Db    1426 ATACACTTTGATGTTACAGTTATGGATTTTGCAAGCCTGTATCCTAGCATAATAAAGGTG 1485

Qy     191 TyrAspVal---PheLysAspValArgMetValPheAlaProProSerSerValGlyLys 209
              ::::::   ::::::   ||||||   |||         |||
Db    1486 CGAAACCTCTCGTACGAGACCGTCAGGTGCGTTCAT------CCCGAATGCAGAAAGAAC 1539

Qy     210 PheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe 229
                    ||||||::::::|||:::        ::::::   |||      |||:::
Db    1540 ACCATCCCCGATACCAACCACTGGGTATGCACGAAAAACAACGGGCTTACATCGATGATA 1599

Qy     230 ArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyr 249
                          |||:::      |||||||||
Db    1600 ---------ATAGGATCGCT-CCGCGACCTGCGCGTCAACTATTACAAGAGCCTCTCAAA 1649

Qy     250 LysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMet 269
           :::|||||||||                    |||      |||      |||
Db    1650 GAGCCAGTCTATAACGGA---------------GGAGCAGCGGCAGCAGTATACTGTGAT 1694

Qy     270 ThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArg 289
                 |||      |||         |||   |||         |||      |||
Db    1695 CAGCCAGGCCCTCAAGGTGGTGCTAAACGCAAGCTACGGGGTGATGGGCGCCGAGAT--- 1751

Qy     290 IleGluAsnGluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLys 309
                                                            :::
Db    1752 ------------------------------------------ATTCCCGCTGTACTTTCT 1769

Qy     310 GluAlaMetSerAlaAspGlnAlaThrArgIleLysTyr------AlaSerLysTyrAla 327
                                      |||      |||      |||      |||
Db    1770 GCCTGCCGCCGAGGCCACCACGGCGGTCGGGCGCTATATCATCATGCAGACCATATCGCA 1829

Qy     328 GlnSerAlaAsnTyrTrpLysAsnSerIle-GlyMetAsnArgGlyLeuAlaArgLeuAs 347
                 |||:::      |||   :::::: |||                        ||
Db    1830 CTGCGAGCAGATGGGCGTAAAGGTGCTGTACGGG-----------------GACACCGA 1871

Qy     347 pValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGl 367
            |    :::       |||      |||||||||         |||   :::
Db    1872 TTCGCTGTTCATAAAGAATCCAGAGGAGCGGCAGATCCATGATATAGTCGAGCACGCCAA 1931

Qy     367 yLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAl 387
            |||         :::|||   |||            :::::::|||   |||:::
Db    1932 AAAGGAG------CACGGC---GTCGAGCTCGAGGTGGACAAAGAGTACAGGTATGTCGT 1982

Qy     387 aLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVa 407
              :::||||||:::                  |||         |||      :::|||
Db    1983 GCTATCTAACAGGAAGAAAAACTATTTCGGGGTGACAAAGTCCGGCAAGGTCGACGTC-- 2040

Qy     407 lArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLe 427

Db    2040 ------------------------------------------------------------ 2040

Qy     427 uLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuPr 447
            |||   |||         |||   |||   :::   |||         :::::::::
Db    2041 -AAGGGCCTGACGGGGAAA---AAGTCGCACACGCCCCCGTTCATAAAGGAGCTGTTCTA 2096

Qy     447 oAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeuProAspIlePheLy 467
             ::::::|||||||||              :::   |||   :::         |||   |||::
Db    2097 TTCGCTGCTCGACATA-----------CTGTCGGCTGTACAGACCGAGGACGAGTTTGA 2144

Qy     467 sAsnVal---------IleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPh 484
             ::::                 |||   |||            |||:::
Db    2145 ATCGGCAAAGCTAAAGATCTCAAAGGCCATAGCGGCATCCGGGAAGAGG----------- 2193

Qy     484 eValPheAspLysSerValValProTyrSerAsp---LysPheHisAlaMetLeuLysSe 503
                 ::::::         ||||||   :::|||      |||:::   |||:::
Db    2194 ----CTGGAGGAGAGGGGGGTCCCGCTGGCGGATCTGGCGTTCAATGTGATGATAAGCAA 2249

Qy     503 rMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLy 523
                     |||   :::   |||   :::   :::
Db    2250 GGCGCCCTCTGAA---TACGTAAAGACCGTCCCGCAGCAC------------------- 2286

Qy     523 sSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIl 543
                 |||   |||||||||   ::::::                  ||||||      ::
Db    2287 -------ATACGGGCGGCCAGACTGCTCGAG--------------AACGCAAGGGAGGT 2324

Qy     543 eGluLysGly 546
             :::||||||
Db    2325 CAAAAAAGGC 2334
```

RESULT 14
US-10-034-623-61
; Sequence 61, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCORP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 61
; LENGTH: 2538
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2538)
US-10-034-623-61

```
Alignment Scores:
Pred. No.:               0.0555      Length:          2538
Score:                   110.00      Matches:         121
Percent Similarity:      37.06%      Conservative:    88
Best Local Similarity:   21.45%      Mismatches:      233
Query Match:             2.96%       Indels:          123
DB:                      9           Gaps:            26

US-10-008-355-2 (1-712) x US-10-034-623-61 (1-2538)

Qy      20 GlyValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlu------------LeuAsn 35
           |||||||||   :::|||         :::::::::::   :::                :::
Db     901 GGCGTGGCGGATTCCGACATACCCCTGTACATGATGCGGGATTCGGCCACGCTCCGGCAC 960

Qy      36 GlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyr 55
              :::||||||      |||                   |||:::               ||||||
Db     961 GGCGTCCATCTGGACCTGTACAGGACC------TTCTCGAACAGGTCGTTCCAGCTGTAT 1014

Qy      56 SerPhe--------------------AspLysProSerIleAlaAsnAlaValValIle 68
           :::|||                              |||::::::   |||::::::
Db    1015 GCATTTGCGGCAAAGTATACAGATTACTCCCTGAACAGCGTGTCCAAGGCGATGCTCGGC 1074
```

```
Qy     69 PheGlyGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
             |||            |||:::::::::  |||      |||     |||   :::
Db   1075 GAGGGCAAGGTCGATTATGGCGTGTCTCTCGGGGATCTCACTCTATACCAGACTGCAAAC 1134

Qy     89 HisCysGlyTyrGlyAla------IleGlnSerGlnSerThrValAspHisAspTyrLeu 106
          :::|||   :::   |||                 ||||||    ::::::    |||
Db   1135 TATTGCTATCATGACGCGCGCCTGACGCTGGAGCTTAGCACCTTTGGGAACGAGATACTG 1194

Qy    107 ArgAspGlyPheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSer--- 125
             |||      |||        :::          :::||||||      :::|||
Db   1195 ATGGACCTCCTGGTGGTGACCAGCAGGATTGCCCGGATGCCCATCGATGATATGTCCCGC 1254

Qy    126 ---------ValLysTyrLeuArgLysIleValLysValThrAspLysValGluGlyGln 142
                            :::::::::|||   ::::::              :::
Db   1255 ATGGGCGTCTCGCAGTGGATAAGGAGCCTGCTGTACTATGAGCACAGGCAGCGCAACGCG 1314

Qy    143 LeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGlu 162
          |||               ||||||:::|||     :::::::::|||:::|||      :::
Db   1315 CTGATACCCCGCAGGGACGAGCTGGAA------AAGAGGTCTCAACAGGTAAGCAACGAC 1368

Qy    163 LeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluPro--------- 179
                |||:::                  |||   :::|||||||||
Db   1369 GCCGTAATCAAGGACAAAAAGTTCCGCGGTGGTCTC---GTAGTCGAGCCTGAAGAGGGC 1425

Qy    180 --------------------------PheTyrSerAsnAsnGluTyrPheLeuIleVal 190
                                    |||   |||               :::   |||
Db   1426 ATACACTTTGATGTTACAGTTATGGATTTTGCAAGCCTGTATCCTAGCATAATAAAGGTG 1485

Qy    191 TyrAspVal---PheLysAspValArgMetValPheAlaProProSerSerValGlyLys 209
             ::::::   ::::::   ||||||   |||          |||
Db   1486 CGAAACCTCTCGTACGAGACCGTCAGGTGCGTTCAT------CCCGAATGCAGAAAGAAC 1539

Qy    210 PheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe 229
                   ||||||::::::|||:::        ::::::   |||      |||:::
Db   1540 ACCATCCCCGATACCAACCACTGGGTATGCACGAAAAACAACGGGCTTACATCGATGATA 1599

Qy    230 ArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyr 249
                     |||:::       |||||||||
Db   1600 ---------ATAGGATCGCT-CCGCGACCTGCGCGTCAACTATTACAAGAGCCTCTCAAA 1649

Qy    250 LysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMet 269
          :::|||||||||                           |||      |||      |||
Db   1650 GAGCCAGTCTATAACGGA---------------GGAGCAGCGGCAGCAGTATACTGTGAT 1694

Qy    270 ThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArg 289
                |||      |||         |||   |||            |||      |||
Db   1695 CAGCCAGGCCCTCAAGGTGGTGCTAAACGCAAGCTACGGGGTGATGGGCGCCGAGAT--- 1751

Qy    290 IleGluAsnGluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLys 309
                                                                :::
Db   1752 -----------------------------------------ATTCCCGCTGTACTTTCT 1769

Qy    310 GluAlaMetSerAlaAspGlnAlaThrArgIleLysTyr------AlaSerLysTyrAla 327
                                     |||      |||      |||         |||
Db   1770 GCCTGCCGCCGAGGCCACCACGGCGGTCGGGCGCTATATCATCATGCAGACCATATCGCA 1829

Qy    328 GlnSerAlaAsnTyrTrpLysAsnSerIle-GlyMetAsnArgGlyLeuAlaArgLeuAs 347
                |||:::      |||   :::::: |||                          ||
Db   1830 CTGCGAGCAGATGGGCGTAAAGGTGCTGTACGGG-------------------GACACCGA 1871

Qy    347 pValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGl 367
           |   :::      |||    ||||||||         |||   :::
Db   1872 TTCGCTGTTCATAAAGAATCCAGAGGAGCGGCAGATCCATGATATAGTCGAGCACGCCAA 1931

Qy    367 yLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAl 387
            |||          :::|||   |||         ::::::|||   |||:::
Db   1932 AAAGGAG------CACGGC---GTCGAGCTCGAGGTGGACAAAGAGTACAGGTATGTCGT 1982

Qy    387 aLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVa 407
             :::||||||:::      |||         |||      |||      :::|||
Db   1983 GCTATCTAACAGGAAGAAAAACTATTTCGGGGTGACAAAGTCCGGCAAGGTCGACGTC-- 2040

Qy    407 lArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLe 427

Db   2040 ------------------------------------------------------------ 2040

Qy    427 uLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuPr 447
            |||   |||         |||   |||   :::   |||         :::::::::
Db   2041 -AAGGGCCTGACGGGGAAA---AAGTCGCACACGCCCCCGTTCATAAAGGAGCTGTTCTA 2096

Qy    447 oAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeuProAspIlePheLy 467
            ::::::|||||||||              :::   |||   :::      |||   |||:::
Db   2097 TTCGCTGCTCGACATA-----------CTGTCGGCTGTACAGACCGAGGACGAGTTTGA 2144

Qy    467 sAsnVal---------IleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPh 484
            ::::               |||   |||                  |||:::
Db   2145 ATCGGCAAAGCTAAAGATCTCAAAGGCCATAGCGGCATCCGGGAAGAGG---------- 2193

Qy    484 eValPheAspLysSerValValProTyrSerAsp---LysPheHisAlaMetLeuLysSe 503
                ::::::         ||||||   :::|||         |||:::   |||:::
Db   2194 ----CTGGAGGAGAGGGGGGTCCCGCTGGCGGATCTGGCGTTCAATGTGATGATAAGCAA 2249

Qy    503 rMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLy 523
                    |||   :::   |||   :::   :::
Db   2250 GGCGCCCTCTGAA---TACGTAAAGACCGTCCCGCAGCAC------------------ 2286

Qy    523 sSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIl 543
                  |||   |||||||||   ::::::                        ||||||      ::
Db   2287 -------ATACGGGCGGCCAGACTGCTCGAG--------------AACGCAAGGGAGGT 2324

Qy    543 eGluLysGly 546
            :::||||||
Db   2325 CAAAAAAGGC 2334
```

RESULT 15
US-10-027-801-61
; Sequence 61, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCORP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 61
;   LENGTH: 2538
;   TYPE: DNA
;   ORGANISM: Cenarchaeum symbiosum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(2538)
US-10-027-801-61

Alignment Scores:

Pred. No.:	0.0555	Length:	2538
Score:	110.00	Matches:	121
Percent Similarity:	37.06%	Conservative:	88
Best Local Similarity:	21.45%	Mismatches:	233
Query Match:	2.96%	Indels:	123
DB:	9	Gaps:	26

US-10-008-355-2 (1-712) x US-10-027-801-61 (1-2538)

```
Qy     20 GlyValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlu------------LeuAsn 35
          |||||||||   :::|||         ::::::::::::   :::               :::
Db    901 GGCGTGGCGGATTCCGACATACCCCTGTACATGATGCGGGATTCGGCCACGCTCCGGCAC 960

Qy     36 GlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyr 55
                :::||||||   |||         |||:::                  ||||||
Db    961 GGCGTCCATCTGGACCTGTACAGGACC------TTCTCGAACAGGTCGTTCCAGCTGTAT 1014
```

```
Qy     56 SerPhe---------------------AspLysProSerIleAlaAsnAlaValValIle 68
          :::|||                              |||::::::   |||::::::
Db   1015 GCATTTGCGGCAAAGTATACAGATTACTCCCTGAACAGCGTGTCCAAGGCGATGCTCGGC 1074

Qy     69 PheGlyGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
             |||            |||::::::::   |||      |||      |||   :::
Db   1075 GAGGGCAAGGTCGATTATGGCGTGTCTCTCGGGGATCTCACTCTATACCAGACTGCAAAC 1134

Qy     89 HisCysGlyTyrGlyAla------IleGlnSerGlnSerThrValAspHisAspTyrLeu 106
          :::|||   :::   |||                  ||||||      ::::::   |||
Db   1135 TATTGCTATCATGACGCGCGCCTGACGCTGGAGCTTAGCACCTTTGGGAACGAGATACTG 1194

Qy    107 ArgAspGlyPheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSer--- 125
             |||      |||      :::            :::||||||      :::|||
Db   1195 ATGGACCTCCTGGTGGTGACCAGCAGGATTGCCCGGATGCCCATCGATGATATGTCCCGC 1254

Qy    126 ---------ValLysTyrLeuArgLysIleValLysValThrAspLysValGluGlyGln 142
                      ::::::::|||   ::::::                  :::
Db   1255 ATGGGCGTCTCGCAGTGGATAAGGAGCCTGCTGTACTATGAGCACAGGCAGCGCAACGCG 1314

Qy    143 LeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGlu 162
          |||            ||||||:::|||      :::::::::|||:::|||      :::
Db   1315 CTGATACCCCGCAGGGACGAGCTGGAA------AAGAGGTCTCAACAGGTAAGCAACGAC 1368

Qy    163 LeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluPro--------- 179
                   |||:::                  |||   :::|||||||||
Db   1369 GCCGTAATCAAGGACAAAAAGTTCCGCGGTGGTCTC---GTAGTCGAGCCTGAAGAGGGC 1425

Qy    180 ---------------------------PheTyrSerAsnAsnGluTyrPheLeuIleVal 190
                                     |||   |||            :::   |||
Db   1426 ATACACTTTGATGTTACAGTTATGGATTTTGCAAGCCTGTATCCTAGCATAATAAAGGTG 1485

Qy    191 TyrAspVal---PheLysAspValArgMetValPheAlaProProSerSerValGlyLys 209
             ::::::   ::::::   ||||||   |||         |||
Db   1486 CGAAACCTCTCGTACGAGACCGTCAGGTGCGTTCAT------CCCGAATGCAGAAAGAAC 1539

Qy    210 PheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe 229
                   ||||||::::::|||:::         ::::::   |||      |||:::
Db   1540 ACCATCCCCGATACCAACCACTGGGTATGCACGAAAAACAACGGGCTTACATCGATGATA 1599

Qy    230 ArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyr 249
                      |||:::      |||||||||
Db   1600 ---------ATAGGATCGCT-CCGCGACCTGCGCGTCAACTATTACAAGAGCCTCTCAAA 1649

Qy    250 LysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMet 269
          :::|||||||||                        |||      |||      |||
Db   1650 GAGCCAGTCTATAACGGA---------------GGAGCAGCGGCAGCAGTATACTGTGAT 1694

Qy    270 ThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArg 289
                |||      |||         |||   |||         |||      |||
Db   1695 CAGCCAGGCCCTCAAGGTGGTGCTAAACGCAAGCTACGGGGTGATGGGCGCCGAGAT--- 1751

Qy    290 IleGluAsnGluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLys 309
                                                             :::
Db   1752 ------------------------------------------ATTCCCGCTGTACTTTCT 1769

Qy    310 GluAlaMetSerAlaAspGlnAlaThrArgIleLysTyr------AlaSerLysTyrAla 327
                                     |||      |||      |||         |||
Db   1770 GCCTGCCGCCGAGGCCACCACGGCGGTCGGGCGCTATATCATCATGCAGACCATATCGCA 1829

Qy    328 GlnSerAlaAsnTyrTrpLysAsnSerIle-GlyMetAsnArgGlyLeuAlaArgLeuAs 347
                |||:::         |||   :::::: |||                        ||
Db   1830 CTGCGAGCAGATGGGCGTAAAGGTGCTGTACGGG-----------------GACACCGA 1871

Qy    347 pValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGl 367
           |   :::      |||      |||||||||            |||   :::
Db   1872 TTCGCTGTTCATAAAGAATCCAGAGGAGCGGCAGATCCATGATATAGTCGAGCACGCCAA 1931

Qy    367 yLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAl 387
           |||         :::|||   |||         ::::::|||   |||:::
Db   1932 AAAGGAG------CACGGC---GTCGAGCTCGAGGTGGACAAAGAGTACAGGTATGTCGT 1982
```

```
Qy    387 aLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVa 407
            :::||||||:::        |||         |||      |||      :::|||
Db   1983 GCTATCTAACAGGAAGAAAAACTATTTCGGGGTGACAAAGTCCGGCAAGGTCGACGTC-- 2040

Qy    407 lArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLe 427

Db   2040 ------------------------------------------------------------ 2040

Qy    427 uLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuPr 447
           |||   |||      |||   |||   :::   |||         :::::::::
Db   2041 -AAGGGCCTGACGGGGAAA---AAGTCGCACACGCCCCCGTTCATAAAGGAGCTGTTCTA 2096

Qy    447 oAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeuProAspIlePheLy 467
           ::::::|||||||||             :::   |||   :::      |||   |||:::
Db   2097 TTCGCTGCTCGACATA-----------CTGTCGGCTGTACAGACCGAGGACGAGTTTGA 2144

Qy    467 sAsnVal---------IleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPh 484
          ::::            |||   |||                  |||:::
Db   2145 ATCGGCAAAGCTAAAGATCTCAAAGGCCATAGCGGCATCCGGGAAGAGG----------- 2193

Qy    484 eValPheAspLysSerValValProTyrSerAsp---LysPheHisAlaMetLeuLysSe 503
                 ::::::      ||||||   :::|||      |||:::   |||:::
Db   2194 ----CTGGAGGAGAGGGGGGTCCCGCTGGCGGATCTGGCGTTCAATGTGATGATAAGCAA 2249

Qy    503 rMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLy 523
                    |||   :::   |||   :::   :::
Db   2250 GGCGCCCTCTGAA---TACGTAAAGACCGTCCCGCAGCAC------------------- 2286

Qy    523 sSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIl 543
                 |||   |||||||||   ::::::                  ||||||      ::
Db   2287 -------ATACGGGCGGCCAGACTGCTCGAG--------------AACGCAAGGGAGGT 2324

Qy    543 eGluLysGly 546
          :::||||||
Db   2325 CAAAAAAGGC 2334
```

Search completed: May 23, 2003, 13:41:19
Job time : 259.87 secs